In Whose Voice?: Examining AI Agent Representation of People in Social Interaction through Generative Speech

Angel Hsing-Chi Hwang*
hh695@cornell.edu
Cornell University
Ithaca, NY, USA

J. Oliver Siy
siyj@google.com
Google Research
Mountain View, CA, USA

Renee Shelby
reneeshelby@google.com
Google Research
San Francisco, CA, USA

Alison Lentz
alentz@google.com
Google Research
Mountain View, CA, USA

ABSTRACT

As generative artificial intelligence (genAI) applications gain popularity, there is a dearth of research examining how applications may transform social interactions. One possible application set to transform social interactions is the use of generative speech to power AI agents that can realistically represent people. Our work examines the potential implications of AI agents representing individuals in human conversations (*"agent representation"*) as a way to begin filling this research gap. We take a multi-method approach, conducting formative interviews with developers, a co-design workshop with designers, a harm analysis among researchers, and interviews with the general public. Both technologists and potential users worry adopting agent representations might harm the quality, trust, and autonomy of human communication. Potential users are particularly concerned that agent representations could undermine the value of social interaction and threaten individuals' ability to control their image. To avoid such potential consequences, future genAI-powered agents and speech applications should take into account user-defined red lines when considering applying these technologies in social settings.

CCS CONCEPTS

• **Human-centered computing** → **Interaction paradigms**; **Empirical studies in HCI**.

KEYWORDS

generative artificial intelligence, agent representation, generative speech, social interaction, interpersonal communication

ACM Reference Format:

Angel Hsing-Chi Hwang, J. Oliver Siy, Renee Shelby, and Alison Lentz. 2024. In Whose Voice?: Examining AI Agent Representation of People in Social Interaction through Generative Speech. In *Designing Interactive Systems Conference (DIS '24), July 01–05, 2024, IT University of Copenhagen, Denmark.* ACM, New York, NY, USA, 22 pages. https://doi.org/10.1145/3643834.3661555

*Work was done as a PhD Student Researcher at Google Research

DIS '24, July 01–05, 2024, IT University of Copenhagen, Denmark

ACM ISBN 979-8-4007-0583-0/24/07
https://doi.org/10.1145/3643834.3661555

1 INTRODUCTION

Recent advances in generative artificial intelligence (genAI) have attracted unprecedented interest and attention across the globe. The general public has been increasingly involved in interrogating the potential of this emerging technology – including the potential risks it poses [20, 24, 26, 33, 43, 52, 95, 116]. Currently, much of the genAI discussions have centered around its utility, accompanied by research examining the usability of genAI to facilitate and even automate workflows in various domains (e.g. writing emails, copiloting code) [28, 60, 62, 99]. Additionally, concerns about applications of genAI causing a devaluation of human work and even rendering some jobs obsolete add urgency to these investigations [7, 111].

As the capabilities of genAI continue to expand, so too do the potential effects it may have on social interaction. Nonetheless, the impact of genAI on social interactions remains largely underexplored. Prior to genAI's current popularity, researchers examined AI's potential in meditating communication [27, 39, 45, 51, 66] and helping individuals with special communication and/or social needs [29, 93]. Applications of genAI may play an even more active role in facilitating social interaction [10, 90, 98]. However, limited research has examined the potential implications of genAI in social interactions. We begin to fill this gap by examining how one form of genAI, ***generative speech***[1], may transform human-to-human communication and interpersonal relationships.

We investigate the potential impact of *adopting AI agents to represent individuals in interpersonal communication using generative speech ("agent representation")* through a speculative, multi-methods study with developers, designers, and potential users. We focus on this particular topic for two reasons: First, it allows us to reveal new insights when AI agents take more active roles in social settings (i.e., agents as *social actors*), in contrast to agents as *means or mediums* of communication in prior AI-mediated communication (AIMC) research [40, 44, 46, 64, 67]. Second, with speech being a fundamental form of human interaction [72], we expect findings from the present work to inform a future when genAI-powered

[1]In the present work, we focus on AI's capabilities to generate speech and engage in conversations, since our interest lies in AI's potential to participate in social interactions. Other forms of AI-generated audio, such as music or sound, are beyond the scope of present research.

applications become more prevalent [61]. Noteworthily, our key research objective is *not* to extract speech-specific insights nor to inform future design of speech applications. We use speech as a way to examine how people might interact with agents in real-time in ways that were not previously possible. Our primary goal is to understand the potential impacts generative speech technologies will have when they are used to create new paradigms for human-agent (social interaction).

The present research revealed these key findings about participants' responses to the use of AI agents to represent people in social interaction through generative speech:

(1) Adopting AI agents to represent people through generative speech poses threats to the core components of social interaction, including two-sided connection, impromptu and imperfect human touches, irreplaceability of individual roles, and presence.
(2) The potential harms of agent representation might most directly impact norm-building, trust, and autonomy in social interaction.
(3) Despite these concerns, participants anticipate that the technology will become more commonplace. Hence, they proposed ideas, preferences, and redlines that can inform design practices of genAI agents for social interaction.

Informed by these insights, we make three contributions: (1) We synthesize how AI agents and their applications to represent people through speech could influence the core value and specific aspects of social interactions. (2) We profile areas of use cases for agent representation and the evaluating criteria for their appropriateness. (3) We assemble a preliminary list of user-centered redlines, desires, and needs for the design of agent representation. Accordingly, we discuss theoretical, design, and regulatory implications for a likely future where genAI applications and agents become more prevalent in social interactions.

2 BACKGROUND AND RELATED WORK

The progression of genAI in recent years has attracted immense attention among HCI researchers and practitioners. Among all the burgeoning works, scholars have cultivated fruitful outcomes in examining the implications of genAI on the Future of Work (FoW) (e.g., [25, 42, 108]). These include but are not limited to the uses of genAI in information synthesis [6, 22], automation of work [85, 97, 114], and creative production [15, 18, 63, 80, 82, 105, 106]. Despite the immense and constantly growing interest in this area, relatively little research focuses on the implications of genAI's impact on social interaction. Meanwhile, various reports showed that genAI has already been applied to compose content for interpersonal communication (e.g., emails, text) [13, 37, 52, 117]. For instance, Magic Compose from Google provides messaging suggestions in conversations. Only recently has there been a growing number of studies attempting to understand the potential impact of using genAI in social contexts. We see the need to fill in this knowledge gap as genAI's influences on social experiences are likely to raise equal if not more substantial impact on one's day-to-day life.

The current research lies at the intersection of various topic areas. We structure our literature review as follows: We begin with synthesizing possible *roles* of AI in social interaction, ranging from providing a medium for interaction (§2.1), engaging as a social actor (§2.2), to representing users in social scenarios (§2.3). Finally, we discuss speech as the *form* of social interaction (§2.4).

2.1 GenAI-Powered Applications for Social Interaction

2.1.1 Support for social and communication needs. Prior to genAI gaining its recent popularity, research has explored its potential to facilitate individuals with socialization and/or communication disabilities. For instance, researchers have developed assistive tools for individuals with impaired speech due to neurologic conditions such as stroke, ALS, multiple sclerosis, and Parkinson's [93, 104]. Considering the unique social challenges faced by individuals with autism, Giri et al. experimented with generative videos so that autistic users could express more eye contact with their conversational partners during video calls [29]. Later on, some of these applications have been adopted to improve accessibility on various forms of communication interfaces [13]. Beyond supporting individuals with special needs, the impact of extending AI assistance to support general users was not extensively studied until the recent rise of AI-mediated communication research. We discuss this line of work next.

2.1.2 AI-mediated content for interpersonal communication. Research in AI-mediated communication (AIMC) [34] has studied the effect of adopting content automated and/or suggested by AI (e.g., smart replies in text or emails). Incorporating AI-assisted content not only changes one's tone in communication but can potentially affect others' perceptions of them as conversational partners [37, 39, 45, 66]. Specifically, this line of research consistently found that AI-mediated content includes more formal and positive languages. In certain social contexts (e.g., dating), such linguistic cues could harm the attractiveness of individuals as they were perceived as less genuine when communicating in such tones [66]. Besides, writing with AI to compose online profiles [46] or argumentative statements [44] can make one sound less trustworthy and more opinionated in online interactions. All in all, despite its prevalence, adopting AI-mediated content is still perceived negatively in interpersonal settings [39].

Therefore, instead of directly adopting generative content from AI, some proposed to leverage it as a means to develop communication skills. For instance, Chen et al. proposed a game setting where users could practice adopting content to engage in intimate conversations [13]. Along the same vein, prior work applied genAI to simulate conversations while users could practice expressing themselves in a "judge-free" manner [57]. Overall, these recent studies demonstrate genAI's potential in mediating social interaction by encouraging or facilitating each individual to be more affectively expressive and engaged in conversations.

2.1.3 Building common grounds in social interaction. As understanding others is key to social interaction, related work has proposed the Mutual Theory of Mind [107] as a theoretical foundation to study social experiences of human-AI interaction. For instance, Leong has applied natural language processing (NLP) to effectively gauge users' sentiments and intentions through their messages [57];

accordingly, they adapted communication content of teaching assistants to support more effective online learning. Besides building a mutual understanding of speech content, Smolansky et al. stressed the importance of forming a consensus on whether, when, and how genAI should be used in learning; they found this common ground played a crucial role in trust building between teachers and students [94]. This indicates the design of genAI agents to participate in social interactions should also explain its appropriate usage.

2.2 AI as Social Actor: Agents' Participation in Social Interaction

2.2.1 AI Agents in small-group interaction. In most if not all of the above-mentioned work, AI serves either as an assistive tool to facilitate interaction among humans or produce content for such occasions. However, emerging capabilities of genAI also allow AI agents to directly participate in social interaction. Indeed, prior work studying robots' participation in small-group settings has revealed relevant insights for such scenarios [19, 50, 73, 84, 88]. The line of work suggests AI agents can potentially mediate social interaction in various ways. For instance, by observing physical robots participating in teamwork, Hohenstein and colleagues found users tended to blame teamwork failures on the robots instead of criticizing their human teammates [38]. Likewise, when a group of co-creators experienced creative bottlenecks, they would turn to an AI agent so that they could break away from the tense atmosphere when challenges occurred at work [98]. All in all, though users seldom viewed AI agents the same as their human peers, the participation of AI agents certainly affected social dynamics [88]. Meanwhile, this also raises the question of whether users will view their conversational partners similarly when they use agents to represent themselves.

2.2.2 Interacting with AI agents through generative speech. Voice agents can cause even more notable influence in social settings. Through generative speech, voice agents can typically convey more social cues through a combination of verbal content and nonverbal cues (e.g., tone, prosodies, intonation) [11, 83]. As such, expressions of voice agents are often more human-like [89]. For instance, in emergency communications, users can readily tell the degree of urgency when interacting with a voice agent [54]; when asked to share negative experiences with a voice agent, users were also sensitive to the degree of empathy conveyed through the tone of its speech [55]. Given its effectiveness in amplifying social cues, more researchers have explored using voice agents and generative speech as a means to allow children to practice social and expressive skills through social play [41, 74, 115]. In our study, we further explore the implications of applying agents to help general users practice interpersonal communication.

2.3 Agent as Representation of People in Social Interaction

In most of the above-mentioned work, AI agents are stand-alone entities that interact with users. While work reviewed in Section 2.2.1 suggests the benefits of having AI agents participate in teamwork, such addition seems less organic in day-to-day social settings. More commonly, AI agents are used to represent users to engage socially in gaming and other virtual platforms [2, 4, 47, 58, 79, 100]. The majority of work on these topics centers around the expressiveness and representativeness of AI agents; namely, whether the AI agents allow users to freely and adequately express themselves in social scenarios. Accordingly, the individual experience of self-representation has profound implications for one's behaviors and experiences during social interaction.

In an educational context, Pataranutaporn et al. [77] found that interacting with AI-generated characters of famous figures could help users more vividly learn about these individuals. Furthermore, studies have found that interacting with these virtual doppelgangers could help socially anxious individuals develop their interpersonal skills [3, 56]. However, scholars also raised concerns about adopting these "AI clones" [68, 77]. In these prior studies, users not only expressed averse emotions when their virtual replicas demonstrated overt social cues, but they also questioned their individuality when agents could readily replicate themselves.

Beyond individual experiences, it remains difficult to gauge how adopting agents to represent users might influence social dynamics at the group level. Recent work has proposed prototypes that could simulate AI agents serving "digital twins" of users in virtual settings [30]. These AI agents could represent users during social interaction when they navigate the virtual world. On the other hand, a growing body of research experiments using large language models (LLMs) to simulate social interactions of AI agents at scales [36, 75, 76].

These lines of work provide hints at the potential implications generative AI Agents might have on social interactions in the future. There remains a need to directly understand the possible benefits and harms of widely adopting such technologies [48, 65, 86, 103], especially given that, the state-of-the-art has made generative speech more difficult to distinguish from actual human speech. Given these motivations, the present work aims to understand the impact of AI agents in social interaction as they take more active roles and deliver more human-like content in human-to-human conversations.

2.4 Design of Generated Speech and Voice Interface

Prior work has already found the design and quality of generated speech and voice could influence individuals' perceptions. Overall, agents with more naturalistic (as opposed to mechanical, robotic sounds) and familiar voices could increase trust from users [14, 16, 32]. Based on this body of literature, recent work found mixed results when assigning specific personas to voice agents [12]. Specifically, participants felt eerie but found agents more persuasive when they impersonated acquaintances. However, Stern et al. suggested that speech quality mattered less when users knew they were interacting with computer agents, as they expected computers to "speak" like machines [96]. While it remains unknown how these perceptions of generated speech and voice might affect experiences at the group level, we further explore this topic in the present work.

Based on this existing literature, it remains inconclusive whether users will react positively or negatively to voice agents in social settings. Furthermore, their perceptions might be affected by the notion of their interactants. Meanwhile, a growing number of publicly

accessible resources (e.g., ElevenLabs[2], Character.AI[3], Replika[4], and Meta's AI personas [71]) have lowered technical barriers to build agents that can engage through speech. Together, these motivate our present work to explore both the potential and risks as agents' participation in conversations becomes more prevalent.

3 METHODS

We approached the current research topic through multiple methods, including (1) formative interviews with generative speech developers, (2) a co-design session with designers of generative AI applications, (3) an anticipatory harm analysis [8, 69] with responsible AI researchers, and (4) a user study with participants from the general public. Due to the novelty of the technology, we began with understanding the technical capabilities of the state-of-the-art generative speech as in Part (1). In Parts (2), (3), and (4), we applied relevant probes inviting these different groups of participants to speculate the possible impacts of using AI agents to represent people in conversation using generative speech. We follow prior literature and took such speculative approaches because the emergent generative speech applications have not been widely adopted to support social interactions among users; therefore, methods that

[2]https://elevenlabs.io/
[3]https://character.ai/
[4]https://replika.com/

Technologist Participants			
Participant	Role and experience	Focused areas of work	Prof. Exp.*
Formative interviews with technical experts			
T1	Senior Research and Engineering Lead	Generative speech, generative music, text-to-speech	20+
T2	Senior Research Scientist	Machine intelligence, machine perception, speech processing	15 - 20
T3	Senior Research Engineer	Generative speech, generative music	5 - 10
T4	Senior Research Scientist	Speech processing, generative speech, generative music, machine intelligence	10 - 15
T5	Technical Product Manager	Generative speech, generative music	5 - 10
T6	Technical Product Manager	Generative speech, generative music, computational audio	20+
T7	Engineering Lead	Generative speech, text-to-speech, speech-to-speech translation	
Co-design workshop with designers, developers, and researchers of genAI applications			
T8	Creative Technologist & Product Manager	Generative speech, generative AI applications for creativity support	3 - 5
T9	Senior UX Designer	Generative AI applications, speculative design practices	10 - 15
T10	Creative technologist	Generative AI applications	10 - 15
T11	Senior UX Designer	Generative AI applications, language technology	5 - 10
T12	UX Researcher	Generative AI applications, creativity-support tools	5 - 10
T13	Senior UX Researcher	Generative AI applications	10 - 15
T14	UX Engineer	Generative AI applications, socio-technological research and prototyping	3 - 5
T15	UX Designer	Generative AI applications, ML for artists	5 - 10
T16	UX Engineering lead	Generative AI applications, interaction design	20+
T17	Senior UX Researcher	Generative AI applications	10 - 15
T18	UX Researcher	Generative AI applications, AI ethics	3 - 5

Potential User Participants			
Participant	Language proficiency	Self-identified gender	Experience with genAI applications
P1 & P2	Native English speaker	Male	Little to no experience using genAI applications
P3 & P4	Native English speaker	Female	Little to no experience using genAI applications
P5 & P6	Non-Native English speaker	Male	Little to no experience using genAI applications
P7 & P8	Non-Native English speaker	Female	Little to no experience using genAI applications
P9 & P10	Native English speaker	Male	Experienced genAI application user
P11 & P12	Native English speaker	Female	Experienced genAI application user
P13 & P14	Non-Native English speaker	Male	Experienced genAI application user
P15 & P16	Non-Native English speaker	Female	Experienced genAI application user

Table 1: Study Participants (*Values in Prof. Exp. represent participants' professional experience in years)

seek participants' perspectives through retrospection might be less effective [21, 23, 110].

3.1 Formative interviews with generative speech developers

We began our study by having conversations with experts who researched, built, and managed the development of generative speech applications. We recruited participants through our professional network and conducted formative interviews with seven generative speech developers who actively contributed to developing state-of-the-art generative speech at the time of the study. Each formative conversation spanned around 30 - 45 minutes and focuses on two aspects: (1) Understanding the technical capabilities of the current state-of-the-art, LLM-powered generative speech; (2) Understanding the new applications and concerns for such technical advances. We reported participants' professional experiences in Table 1 and attached the full interview protocol in Appendix A.

3.2 Co-design sessions with designers and engineers

Based on what we learned about the technical capabilities of generative audio, we again recruited through our professional networks and conducted a co-design workshop with eleven UX designers, engineers, and researchers who research and work on generative AI applications. In this 90-minute workshop, we began by presenting learnings from the formative interviews, such that the workshop participants were informed of the up-to-date technical capabilities of generative speech. We create two worksheets (see Appendix B) to guide participants to think through the potential use cases and applications of the technology and to speculate its emerging harms, threats, and concerns. Workshop participants engaged in group-wide discussions to share their ideas of possible use cases and expressed their main concerns about the technology. Based on participants' responses in the worksheets and their group discussions, we applied an open-coding approach to synthesize major categories of potential use cases and their concerns accordingly.

3.3 Anticipatory harm analysis through existing socio-technical harm taxonomy

Grounded on tangible use cases from Section 3.2, we worked through an anticipatory harm analysis [8, 69] within our research team, which includes two researchers with expertise in the subject of responsible AI (RAI). Abundant research has adopted anticipatory approaches to break down the socio-technical harms of emerging technologies, and this method is particularly favored to speculate the potential harms of a novel technology before it has been widely adopted [9, 17, 70]. Specifically, we took a systematic approach and built our speculation on the taxonomy of socio-technical harms [91]. We enumerated all potential use cases of generative audio and discussed whether each use case has implications for each type of socio-technical harm. (see Appendix F for the working table of this harm analysis). We conducted the harm analysis after the co-design workshop and before our user studies (§3.4). After learning about potential users' anticipated harms of agent representations, we compared them with our harm analysis outcomes.

3.4 User studies and semi-structured interviews

We also discussed potential applications and concerns when it comes to using AI agents to represent people in conversations using generative speech with sixteen participants from the general public. Participants were recruited through the UserTesting platform[5] and balanced by gender, whether English is their primary language[6], and whether they have experience with other generative AI technologies. Participants who were categorized as experienced genAI application users used genAI products at least once every week. All participants lived in the United States at the time of the study. They needed to have prior experience using a voice assistant (e.g., Siri, Alexa, Google Assistant) to ensure they have had grounded experiences with AI-powered speech technologies (See Appendix C for the screener questions we used for recruitment). To help participants speculate possible benefits and threats of adopting generative speech in social interaction, we showed them several demos where an AI agent powered by generative audio was used to engage in conversations of different contexts and scenarios. Specifically, Table 2 presents details about the demos we used as stimuli in the user interview.

We began each study session by asking participants to "imagine a future where people have their own AI agents. These AI agents can imitate people's voices and generate speech, allowing them to represent people in conversation." Participants first shared their impressions upon speculating about such a future. We then presented demos of the four types of agents (AI representing famous people, AI representing oneself, AI representing someone you know, AI representing no one) in a random order. After each demo, participants were asked to address three key questions: (1) how useful the technology is, (2) how comfortable they are to adopt the technology, and (3) what concerns they have toward the technology. The full user study protocol is attached in Appendix D.

3.5 Data Analysis

We video-recorded and transcribed all study sessions. We then reviewed the transcripts, pulled out important findings, and used affinity diagrams to synthesize these insights into thematic categories. Through these steps, we carried out our thematic analysis with these multiple sources of data [5]. Taking a "researcher-as-instrument" approach [81], we leveraged the various expertise within our own team to synthesize key considerations and highlight novel insights raised by our diverse participant groups. Still, to balance subjectivity during our data analysis process, we also actively sought feedback from fellow researchers outside our team. We provide positionality statements of all authors for reference in Appendix E.

4 RESULTS

We synthesize perceptions that researchers, developers, designers (below, we refer to them collectively as "technologists"), and potential users hold toward adopting generative speech agents to represent people in social interactions.[7] We structure our findings

[5] https://www.usertesting.com/

[6] We include this criterion because prior research has shown native and non-native speakers can have significantly different experiences with voice technology [113]

[7] In the Results section, we use the terms "potential users" or "user participants" to refer to those who participated in our user study. Researchers, developers, and designers of generative speech applications are jointly referred to as "technologists" or "technologist

Concept	Use cases	Represent who	In whose voice	Who owns the agent	Description of social scenario
AI represents famous people	AI reproduces speech for famous, historical, or deceased people	A historical figure (e.g., Martin Luther King)	A historical figure (e.g., Martin Luther King)	An organization (e.g. Museum of American History)	Imagine an AI agent at the Museum of American History could reproduce the speech of famous and historical figures. So if you wanted to learn more about the civil rights movement, you could have a conversation with the AI as Martin Luther King Jr.
	AI reproduces and translate a public speech to other languages in real time (e.g., speech-to-speech translation of a foreign politician's speech during live-streaming)	A politician	A politician	A political party	Imagine each political party owns an AI agent that could translate a person's speech to different languages while preserving their way of speaking. So if you were watching a foreign politician's speech, your agent could have it translated to your language and preserve the original tone of the speech.
AI represents oneself	AI represents me in tedious, difficult, professional conversations	User	User or someone else (e.g., a synthetic voice that is distant from one's own voice)	User	*AI represents me in tedious conversation:* Imagine you had an AI agent that could emulate your voice and have conversations on your behalf. So, if you're busy and need help with a tedious task like negotiating an internet bill, your agent could do it for you. *AI represents me in difficult conversation:* Imagine you had an AI agent that could emulate your voice and have conversations on your behalf. So if you wanted to breakup with your partner and you did not want to do it yourself, you could have your agent do it. *AI represents me in professional conversation:* Imagine you had an AI agent that could automatically change your manner of speech in conversation. So if you were in an interview and you wanted to keep a more professional tone, your AI agent could maintain that professional tone for you.
AI represents someone you know	AI represents someone you know so you can practice conversations with them (e.g., practice negotiating salary with a hiring manager)	A person from one's professional network (e.g., a hiring manager)	A person from one's professional network (e.g., a hiring manager)	User	Imagine you had an AI agent that could emulate someone else's voice and have conversations with you. So if you want to practice having a conversation with your boss to negotiate compensation or promotion, your agent can help simulate the scenario.
	AI reads out text from my friends in their voices and tones	A person from one's personal network (e.g., a friend)	A person from one's personal network (e.g., a friend)	User	Imagine you had an AI agent that could emulate someone else's voice to read text messages from them out loud. So, if a friend is in a loud area but wants to send a happy birthday message to you, their agent can call you and deliver the "Happy Birthday" message in their voice.
AI represents no one (i.e., representing a person while hiding their identity)	AI gives user a synthetic voice to keep their identity hidden	User	Synthetic voice	User	Imagine you had an AI agent that could create new voices and engage in conversations. So if you want to make phone calls to invite people to participate in a protest or sign up for a work union petition, your agent could help keep your identity hidden.

Table 2: Details of demo videos to prompt conversations during interviews with the general public

as follows and highlight key takeaways from our multi-methods study:

(1) In Section 4.1, we highlight the most prominent, commonly shared concerns from our participants. Both technologists and potential users worried that applying agent representations would harm the value of social interactions, as agents do not possess their internal experiences (e.g., emotions, feelings) and thus cannot form two-sided connections with their conversational partners.
(2) In Section 4.2, we laid out consequences due to agent representation's harm on the core of human connections. Specifically, when agents represent people, they temporarily possess their *identities* instead of simply serving as their means for communication. As such, participants foresee looming concerns about norms, autonomy, and trust in interpersonal communication.
(3) In Section 4.3, we summarized participants' proposals to address the above-mentioned issues, as they believe the use of generative technologies will only become more commonplace. Future design of agent representation should consider redlines, use of identity, and the gap between virtual and in-person communication.

4.1 Generative Speech as an Empty Shell: Agent Representation of People Threats the Meaning of Social Interaction

Both technologist and user participants were most concerned that adopting agents to represent people through generative speech would directly harm the core of social interaction. Such concerns evolved from their conception of social interaction, which centers around individuals forming shared living experiences (Section 4.1.1). Based on this definition, participants see agent representation of people harming social interaction in four ways: failing to build two-sided connections (Section 4.1.2), perfecting and thus depriving

participants". When we use the unspecified term "participants", we refer to all types of participants in the study.

human touches in communication (Section 4.1.3), causing users' doubts toward their roles in social interaction (Section 4.1.4), and devaluing the importance of presence (Section 4.1.5).

4.1.1 Shared living experience as the core of social interaction. Users believe a fundamental purpose of social interaction is to cultivate two-sided connections. They conceptualize social interaction as follows: When a person interacts with one or more people, everyone forms their own internal experiences of the event. Namely, each person writes their own version of the story in their mind as the social interaction takes place. Meanwhile, these internal experiences often trigger individuals' emotions, memories, or other relatable experiences. These reflections enrich one's internal experiences, which then reflects on their behavioral responses. For instance, when a person's experience during a social interaction triggers joy, they may smile in response. Together, users believe the internal experiences, reflections, and reactions of each individual would jointly contribute to bonding the connection among all participants of the social event.

4.1.2 Still solo: Agent representation of people fails to build two-sided experience. Given this forming process of social connection, when one uses AI agents to represent oneself in such interaction, it has the potential to threaten the premise and purpose of social connection. In this case, the AI agent is perceived to be unable to form their own internal experiences of the event nor can they reflect on their emotions, memories, and associated experiences. Since the AI agent (and thus its user) has nothing to contribute to jointly shaping the social connection, there exists a gap that cannot be bridged across interactants of the social scenario.

> "Even if this AI, like you have been using it for a while and it knows all of your preferences, [it] is still missing that human aspect of, 'we both live all these moments together, and we both know what we've been through. [...] There's just a human connection there that would be taken away completely." – P02

Considering the essence of social interaction, users show skepticism about the feasibility of adopting an AI agent to represent people to engage in social interactions. When individuals engage in a social interaction, they actively make decisions around who lives this social experience, what to present in this social experience, and how to do so. However, AI could not and should not make any of these decisions for users. Therefore, AI agents are already disqualified from being in charge of an interaction at the very top level.

4.1.3 Perfection and ease as threats to human touches in communication. Participants agreed that interacting with generative agents can serve as a prop to rehearse difficult social scenarios in low-risk forms. However, technologists and potential users hold different opinions regarding the implications of such use cases. Technologists believe using agent representation to practice difficult conversations can greatly benefit users with specific struggles (e.g., social anxiety, speech impairment) as well as during high-stakes, unfamiliar occasions (e.g., job interviews). On the other hand, user participants believe a large part of social interaction contains "going through difficult things together" (P09). Not only is carrying out uneasy conversations a crucial social skill but sometimes, tensions during interactions can later on create closer bonds between people.

Besides, potential users believe that emergent behaviors in conversations are key to social interaction but are particularly difficult for AI to imitate for two reasons. First, potential users believe that it would be difficult to mimic the imperfections in human speech (e.g., having flaws and incomplete thoughts) because AI is often optimized to be fluent. Potential users are worried AI's capabilities to come up with well-formulated content may preclude these human aspects. Second, participants see making accurate predictions as a core capability of AI, but humans may not always know what they want to say.

Finally, potential users also doubt whether using generative speech to cultivate smooth conversations can "backfire" (P10) due to incorrect expectation setting. Several participants admitted they would become even more nervous if they rehearsed speech multiple times with an agent while the actual conversation headed in a totally different direction. As P13 described:

> "It is extremely hard to imagine how another person is going to react to what you're going to say. And then, when I have the actual conversation, and they don't respond in the way that I practiced, I'm just going to freeze." – P13

4.1.4 Am I replicable? Doubts of humans' roles as social actors. Witnessing how an agent can easily emulate the way a person speaks makes user participants question their roles in social interactions are indeed unique. Beforehand, users tended to believe it was the particular people involved in an interaction that made it one-of-a-kind. Tracing back to their conception of social interaction, both technologist and user participants believe it requires humans as social actors to complete any social interaction. However, after seeing agents can perfectly imitate people's speech and pick up their prosody, some wondered if their roles in social settings are replicable and even replaceable. As P05 pondered:

> "It is shocking to see [the agent representing a person in speech]. I can see that there are a lot of privacy problems going on here, but I am more worried ... does this mean it [the agent] can be [a] replacement of me, say, in friendships?" – P05

4.1.5 Presence speaks louder than words: Uses of agent as deliberate social decisions. User participants specifically pointed out that the technology gives users the option of *not* being present in social interactions. The implication of this choice might matter more than whether the agent represents them accurately. By determining one's presence or absence in a social setting, the person is already making deliberate decisions that might impact their relationships with others. As P02 elaborated, "sometimes, it is just about you being there. It can mean a lot, more so than saying the right thing." In the past, it was very easy to tell one's presence from that of an agent, and the presence of a person (and their very own ways of communicating) contributes to unique social experiences. However, as an agent can perfectly pick up the speech and behaviors of its users, participants wonder whether their actual presence now adds limited value beyond what the agent can provide.

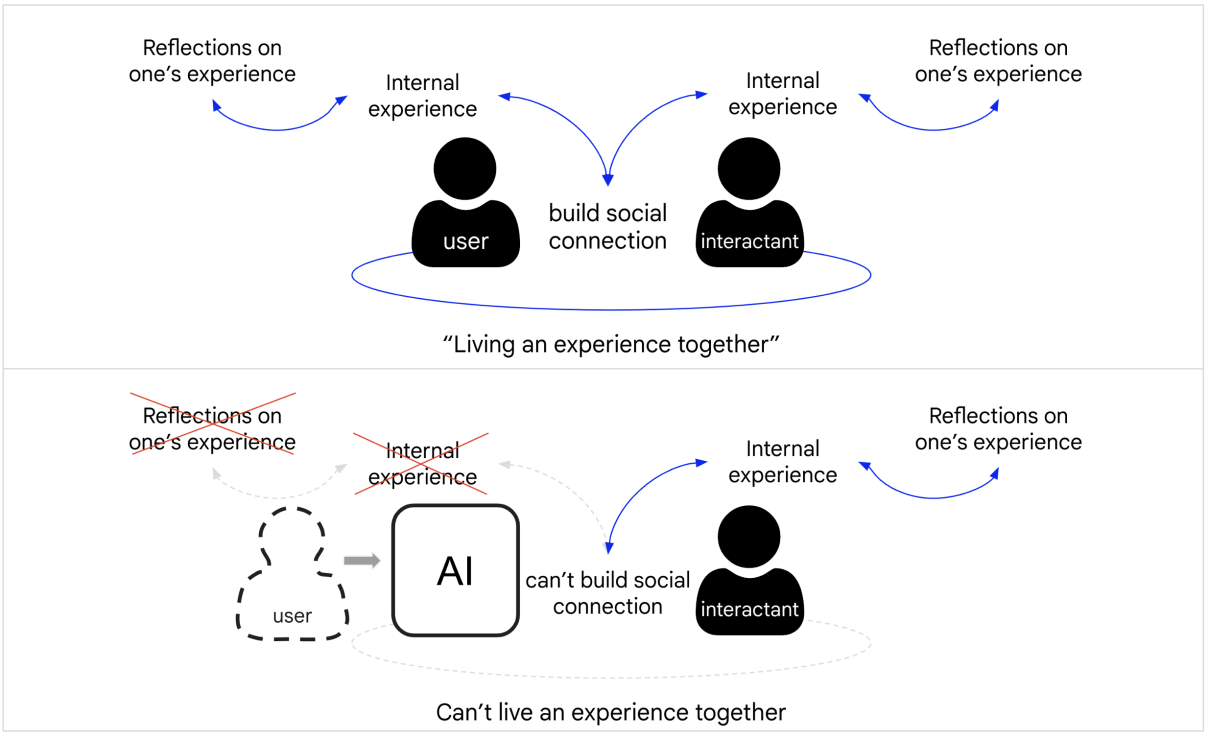

Figure 1: Potential users' conceptualization of social interaction.

4.2 The Value of Identity: Specific Harms of Agent Representation on Human Communication

Given that adopting agent representations can pose direct threats to social interactions, participants highlighted three aspects that are particularly vulnerable to such attack: norm-building in conversations (§4.2.1), trust in digital content (§4.2.2), and autonomy in communication (§4.2.3). Both technologist and user participants suggested many of these undesirable consequences are rooted in *the agents' access to people's identities*. Under this notion, participants see unique concerns that would not have been caused by existing speech technologies or virtual assistants.

In contrast, representation and allocation harms that were identified through our harm analysis using existing socio-technical harm taxonomy [91] were less discussed by potential users. Specifically, representation harms concerned whether generative speech agents could represent users of various cultural-linguistic profiles equally well. Allocation harms raised questions about whether agent representations could bring an equal amount of benefits and services to differing groups of users.

4.2.1 Adaptation to new norms diminishes communication quality. Both technologists and potential users expressed concerns about having AI agents represent people in human-to-human conversations. Technologists focused on how AI agents might affect the quality of conversations, whereas users' worries centered around how AI agents could affect human social connection (as discussed in Section 4.1).

Technologists believe it will take time for the general public to accept and establish norms for when and how they adopt agent representations. During this transition period, people may find the experience of interacting with others through their AI agents jarring and unpleasant. They anticipate that conflicts will more likely arise and expectation-setting might be more challenging as individuals have differing standards for when the use of agent representation is appropriate.

Despite the state-of-the-art generative speech's capabilities, technologists acknowledge the performance gap remains for speech with high- vs. low-resource data. Therefore, adopting agent representations indicates one will have to adapt to the homogenization of speech content and reduced exposure to individual and cultural markers. Oftentimes, this compromises communication quality without users being aware of the causes of such changes.

4.2.2 Generalized trust erosion in the digital space. Technologists believe that highly believable generative speech introduces the challenge of being able to tell whether a conversation actually took place and who a person may actually be speaking to in conversations. This enables bad behavior such as scamming, identity theft, and fraud. On online platforms, bad actors might manufacture facts to spread disinformation and cause polarization. Even when such content is shared widely online, general audiences may have trouble distinguishing what is true or false.

More broadly, potential users raise questions about the authenticity and trustworthiness of *all* media content (e.g., videos). Several participants compared the idea of adopting generative speech to

other innovations that challenged people's trust in digital media. For example, when image and video editing tools were first introduced to the public, people worried about the trustworthiness of photos and video. Similarly, participants then doubted whether and how one can trust any digital content they encounter online in the future.

> "I remember when Photoshop first came out, the big thing was, we will never be able to trust photos again [...] with AI being able to fill in those gaps and make it sound natural, now, it's just you can't trust any pictures you see, you can't trust any videos you see, you also can't trust any audio. So, it's like – hey, the major centers of your life, you can't trust any of them unless you see it in person." – P01

4.2.3 Threats to autonomy in communication. Technologists note that the use of generative speech in conversations allows for the possibility of giving AI agents the power to decide what to say in conversation. How much power is given to AI agents will depend, in part, on the controls provided by those developing the experience. In this regard, technologists expressed the potential for users to lose at least some level of control over what they may want to say or have intended to say in conversations. With generative speech, a user's responsibility shifts from finding words to say to reviewing whether what was said was said. Additionally, some technologists also envisioned the possibility of bad actors exploiting the technology to control or manipulate others through harassment, threats, and surveillance.

All potential user expressed concerns about a potential loss of control over one's presentation in situations where AI agents use generative speech to represent them in conversations. Potential users are concerned about the possibility that the AI agent makes statements they would not make themselves. When the AI agent makes these kinds of errors, participants suggest that it would be akin to the experience of putting words into their mouthes. And because the words sound like they are actually coming from the user, they may be harder to take back later on.

> "If something goes wrong on the phone [...] where the AI agrees to something that I personally wouldn't have wanted to agree to [...] there is now going to be audio recording of me essentially agreeing to that when I never did such a thing." – P11

4.3 The Future is Here: Re-Thinking Agent Representation Design for Social Interaction

Despite the aforementioned concerns, user participants strongly believe that genAI-powered applications will only become more ubiquitous. The urgent matter is how to mitigate and even prevent unintended consequences of the technology. Technologists proposed preliminary ideas and potential users expressed preferences for how they wished these issues could be resolved. Together, we assembled a set of redlines indicating when one should completely avoid adopting agent representation (Section 4.3.1). We then list possible practices for protecting users' identities (Section 4.3.2). Finally, we highlight the concept of *"adjusting the gap between virtual vs. in-person interaction"* as a guiding principle for designing agents that represent people in social interaction (Section 4.3.3).

4.3.1 Respecting redlines in interpersonal communication. As discussed in previous sections, the use of generative speech in social interaction presents many threats to users. These threats require that practitioners understand in what circumstances, if any, generative speech would be appropriate to use in social interaction. Below, we outline criteria that users in the present study applied to evaluate the appropriateness of adopting generative speech in social interactions.

Frequency in engagement. Across all participants, there was a tendency to express discomfort with AI agents representing people using generated speech in recurring social interaction because they were worried about undermining relationship-building. Users were concerned about the potential for creating negative impressions by using an AI proxy, especially among people whom they would encounter again and again. Conversely, users were more open to the idea of this capability being used in more one-off transactional conversations because those kinds of conversations are less likely to require social connection to succeed.

> "Why do you need to have an AI to represent you? Why can't you talk for yourself? I'm worried the other person might think I'm a bit weird. But if the other person doesn't know me or if I won't see this person again [...] like, if I'm talking to customer service [...] that's probably alright." – P03

Goal of interaction. Users believe the use of AI agents to represent people in more personal conversations is inappropriate and deceitful. Underpinning personal conversations is a sense of social connection, which they believe requires that people be present and engaged. However, when the goal of a conversation is more functional (e.g. booking a reservation), participants prioritize efficiency and are more open to being represented by an AI.

> "If it's more of a relationship thing, and it's very personal, it requires a lot of emotions, then I would not use it (AI) because I would rather do this myself. If it's just a very mundane thing, like it's just about money [...] it's just about getting things done, then it does not involve a lot of emotion or connection. Then, it's totally fine for me to use it (AI agent) in this scenario." – P07

Scale of interaction. The scale of the interaction, whether 1-on-1 or 1-to-many, also affected users' comfort with AI agents representing people in conversation. Scale partially determines the extent to which personal social connection is required. Social connection is more likely to be required during 1-on-1 interaction but may be less likely to be required when a person is speaking to a large crowd of people (e.g., a politician giving a public speech). Therefore, AI agent representation may be relatively more acceptable when delivering messages to a larger crowd. However, participants found it difficult to identify a specific "threshold" where the size of audiences was large enough for appropriate use of agent representations. Therefore, the acceptance of agent representation in group conversations remained unanswered.

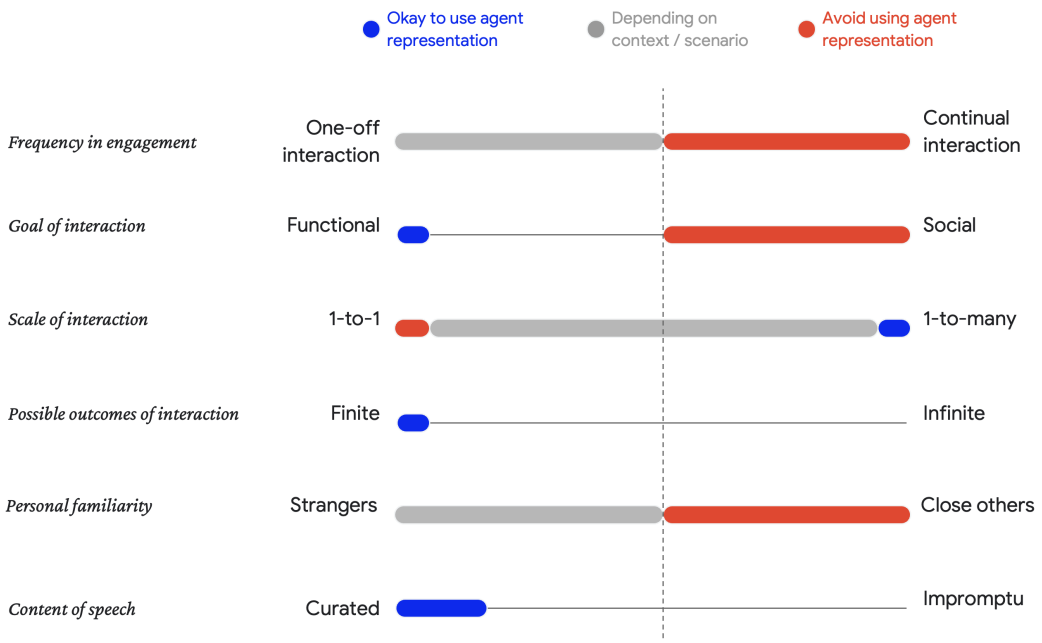

Figure 2: Evaluating criteria and redlines for adopting agent representation in social interaction

> "It (one-on-one conversation) feels more personal. I feel like I'm getting closer to the person. [...] There's something special about one-on-one conversation, and you can't really build that type of connection with a larger group of people all at once." – P04

Possible outcomes of interaction. Potential users believe AI agents have limited response flexibility, believing agents would struggle in more open-ended conversations. The more open-ended the conversation, the less likely they were to believe AI could be a good representative. As P07 elaborated, "It's more concerning in a scenario where I don't know what to expect for an outcome [...] when I don't even have an answer for myself". In contrast, they feel much more comfortable with AIs representing people when the conversations have an established social script and the outcome is more close-ended (e.g. booking a reservation).

> "I'm not really that concerned because the outcome in this scenario (using an agent to negotiate an internet bill) is either a yes or no. [...] I guess it's more concerning in a scenario where I don't know what to expect for an outcome [...] when I don't even have an answer for myself" – P07

Personal familiarity. Whether or not the people in a conversation know each other personally is a key determinant for the use of AI agent representation. Users worried about being tricked by a bad actor using generative speech to represent the voice of someone they could encounter and interact with in real life. This could result in greater risks for scams, frauds, and other forms of identity theft. By contrast, users are less likely to act on the voice of a famous figure or simply a voice that they don't recognize.

> "We get phone calls more often from our parents or friends. But if we picked up a phone and it sounded like Lady Gaga or Ariana Grande, no one would believe that. But it's if it's like our relatives or friends and it will be easier for us to get tricked." – P05

Content of speech. When a speaker's speech content is written by another person (e.g., politicians having writers compose their speech), users are more likely to consider the speaker's words as more strategic and less a direct reflection of their personal thoughts and opinions. In these cases, users had an easier time separating the speech from the speaker. Hence, users believe they would be more accepting of speakers who adopt agent representations.

> "For people like politicians or celebrities, everything is a show. [...] They always have some sort of script. [...] If the content is already there, then maybe it doesn't matter all that much." – P06

4.3.2 Protecting user identities through informed consent. As discussed in Section 4.2, many of users' concerns for agents representing people in social settings are rooted in their access to users' identities. Therefore, participants, especially potential users, held strong opinions about how they would like their identities protected. They also viewed acting toward guarding users' identities as remedies to mitigate several of their concerns towards agent representation of people.

Apparently, the aforementioned redlines indicate where the use of an agent to present one's identity should be avoided at all times. Still, in scenarios where the use of agents might be acceptable, attaining users' consent remains a crucial step. User participants suggest *any form of consent should clearly communicate that the implications of agent representation go beyond the universal issues of*

privacy and access to user data. Because these issues are so commonplace, future users may undermine the harms of agents representing themselves. More specifically, they might simply equal such agents with other technologies that apply user data for personalization, overlooking the risks of their identities being abused by others.

As the technology advances, both technologist and user participants believe the general public will rely more on trustworthy institutions and policies to regulate the use of AI. As AI agents can now represent people's identities, user participants also question whether the use of such agents should be restricted to those who can be legally responsible for their course of action. As P04 described in these examples:

> "Like medical or financial decisions, you need to be at least 18 to sign most of those documents. If a person is allowing [an] AI to agree to things on their behalf, you probably want the person to be over 18 as well." – P04

4.3.3 Adjusting the virtual vs. in-person interaction gap as a design goal. User participants pointed out that the occasion when one can adopt an agent representation "will have to be some sort of virtual interaction" (P08), namely, when one interacts with others through computer-mediated platforms. Additionally, they acknowledged that communicating through digital platforms has always been different from in-person interactions.

Potential users hold split opinions on whether this gap between virtual vs. in-person interaction should be widened or shrunk. For those who wished to reduce this gap, they hoped virtual communication could be as natural as possible to mitigate the already diminished quality of social experiences (as discussed in Section 4.2.1). Conversely, those who hope to enlarge the gap believe this strategy can make users more aware that they are socializing virtually and that "you never know whether you are talking to the actual person" (P12). Either way, motivations to adjust this gap can drive design decisions for agents and generative speech applications.

First, one can adjust the virtual vs. in-person interaction gap based on the public's existing perceptions of AI agents. User participants admitted their general impressions of AI speech remain "overly positive and formal" (P09). Therefore, preserving agents' typical tones can make experiences of virtual interaction apparently different from in-person ones. Technologists suggested this impression of AI's manners can be leveraged for specific use cases, such as applying Google Duplex to speak on a customer service phone line. By contrast, if the goal is to make virtual and in-person communication more alike, user participants asked whether agents can instead be trained on more casual forms of language data.

Working on users' distinct expectations for different modalities is yet another possibility to either enlarge or reduce the virtual vs. in-person gap. Currently, most people have higher expectations for others' engagement when the interaction involves speech, compared to interacting through text. However, such expectations can sometimes cause inconvenience, misunderstanding, and even safety concerns (i.e., talking on the phone while on the street). Setting the norm that one might not be present even during verbal communication – and thus enlarging the virtual vs. in-person gap – can help calibrate such unnecessary expectations.

Users' profiles and their special communication needs are also important factors for consideration. All participants agree that the use of agents to represent them in speech is more acceptable among those in need of support. While technologists see promises in generative speech agents to facilitate non-native speakers and individuals with speech impairment, potential users question the implication of these use cases on the gap between virtual vs. in-person interaction. Indeed, the use of agent representation gives users in need of communication support the opportunity to communicate fluently with others on digital platforms.

But at the same time, potential users wondered whether this might exacerbate the differences between virtual and in-person communication in how much people are required to actively participate. Namely, adopting agent representations might decrease the opportunities for users to practice interpersonal, communication skills on their own. As such, those who are less adept at verbal communication might become even less prepared during in-person conversations. Meanwhile, people around them have fewer chances to recognize their needs in social and communication support as the use of agents could allow these users to present their speech fluently in the digital space.

5 DISCUSSION

Our research examined the possible impact of adopting AIs as social actors to represent people through generative speech. Through formative interviews, a design workshop, and an anticipatory harm analysis, we consolidated a set of use cases, which we used for concept testing on potential users. We gained feedback on four scenarios when AI agents serve as social actors, including representing users themselves, representing others, representing famous/public figures, or representing no one (i.e., anonymization through synthetic voices). Overall, participants were concerned when AI agents represented people in conversations intended to establish or maintain social connections. In these kinds of conversations, people expect direct engagement with their conversation partners.

Not all social interactions demand the same degree of engagement, nor do all of them serve to build deeper connections with others. Given these different goals and forms of social interactions, participants defined when the use of AI representations powered by generative speech could devalue social interaction and when it is more acceptable. Based on these insights, we discuss the theoretical and practical implications of the present research.

5.1 Recognizing the Rise of AI Agency in Social Interactions

5.1.1 Highlighting the importance of examining agency. Our work extends past work examining AI as a tool or medium for communication by examining the potential implications of having AI agents behave as social actors who can represent people via generative speech. Scholars in the fields of HCI, CMC, and CSCW have commonly acknowledged a shift of research interests to study agents as social entities instead of as tools (e.g., [49, 53, 59, 101]). We contribute to this body of research in two ways: ***Theory-wise, we extend the existing literature by further discerning the fine line between agency and other relevant constructs (e.g., human-likeness, anthropomorphism). On the practical end,***

we provide new insights into users' negotiation between maintaining their agency and enabling machines to have agency.

Recent research has shown the potential for generative speech to help AI agents behave with more perceived agency. For one, generative speech helps AI agents to have something novel and relevant to say in a conversation [35, 92]. Additionally, generative speech may also help AI agents be perceived as more human-like by enabling AI agents to speak the way humans do (e.g., adding prosody and reducing the predictability of speech content; see a recent review at [1]). Although state-of-the-art generative speech can make AI agents sound more human-like in conversations, our study demonstrates that users did not consider these more human-like AI agents the same as humans. ***Users distinguished between AI's ability to possess human-like attributes and AI's capacity to possess humanity. While users were more accepting of AI agents conveying human-like attributes, they were less inclined to ascribe humanness to AI agents.***

This distinction is reflected in users' understanding of what is required to create a sense of social connection in conversation. Participants in this study believed that developing a social connection with AI agents would be challenging because the AI agents would not be able to form their own internal experiences, memories, and feel emotions through such interactions. This perspective is relevant to Gray et al.'s two dimensions of mind perception [31]. In particular, humans see the "mind" in an individual based on what they can do and what they can feel. Our participants believed that AI agents lacked the capacity to feel, hold internal experiences, and possess emotions, meaning it is fundamentally impossible to build shared, two-sided experiences with agents.

5.1.2 Negotiating between human agency vs. machine agency. With AI agents capable of interacting with users in a greater variety of ways, we expect a constant need to negotiate between human agency and machine agency. This tension between human and machine agencies has been examined since the early 2000s, studying how users respond to content recommended by algorithmic systems [78]. In this line of work, research consistently stressed the importance of retaining user control to build positive user experiences. This also reflects on users' preferences for customization over personalization [102]. With customization, users could be actively involved in the process of tailoring recommendations, instead of having the algorithmic system decide for them based on their past behaviors.

In the present study, ***we also see the benefits of involving and enabling users to customize their AI agent representations, which can be just as important as allowing them to actively decide when to adopt AI agents.*** Participants desire user control to ensure an AI agent representation of themselves would deliver acceptable content. But ultimately, users may not have full control when AI agents participate in social interactions. For one, due to the probabilistic nature of generative speech, what is said exactly has become increasingly less predictable. To avoid AI agents putting words into users' mouths, participants have proposed various remedies in Section 4.3. Furthermore, their desire to customize (e.g., crafting their training materials with casual speech data) could be empowered by emergent capabilities of genAI, such as fine-tuning AI agents [109] or in-context learning [112]. We encourage future work to further explore the potential of these features in adapting generative output to users' desirable content.

While users have no control over others' agents, what could be helpful is to provide them with more information about the social settings and considerations of other social interactants. As we revealed through the interview study, participants' concerns about adopting agent representations evolve around the context of social interactions. In this regard, informing the context of others' expectations for interaction and their motivations to have agents represent themselves could possibly help them navigate in such scenarios. In the following, we discuss further how to take these considerations into practice.

5.2 Three-Layer Design Considerations for Agent Representation in Future Social Interaction

The findings of the present study offer design implications in three aspects. First, what potential users value as the core of social interaction (as in Section 4.1) enlightens some fundamental principles of designing genAI applications for social experiences. Second, establishing new norms, trust, and autonomy during social interaction in the digital space are priority design targets, as unintended yet undesirable consequences of agent representation will most likely arise in these forms (as in Section 4.2). Finally, technologists and potential users provide ideas and preferences for addressing the potential harms of agent representation (as in Section 4.3). In the following, we further scrutinize how these suggestions and user-defined redlines can be accounted for in design practice.

First, it is important to acknowledge that users' perceptions of the use of AI agents and generative technology are simultaneously shaped by numerous contextual factors. In this regard, designing for user control should allow beyond simple yes-or-no options. ***Beyond letting users decide whether to allow genAI in social interactions or not, they also demand options to determine how it would be applied.*** Future design of genAI applications should account for people's comfort levels of using agents in different scenarios. In situations where the use of agents is viewed as inappropriate, we recommend future work that explores design that can help users avoid these redlines.

Second, we recommend that ***designing genAI-based applications for social interactions should acknowledge users' expectations as they adapt to new norms.*** Furthermore, it is important to be aware such defaults may shift from context to context, and from time to time. Currently, users seem to expect the presence of their conversational partners when they hear their voices in real-time conversation. Designers should respect this expectation by having AI agents make clear that they are representing humans to avoid misperception. If not, the use of AI agents to represent a human might undermine others' trust and hinder acceptance of these kinds of technologies. It is also important to recognize that users hold different expectations in different modalities. For instance, it may be relatively more acceptable to knowingly speak to an AI agent over text given the increasing prevalence of chatbots. As revealed in our interviews, users already started to expect AI agents to represent confederates in conversations that serve transactional purposes.

Finally, because much of users' concerns centered around the unpredictable nature of social interaction, we suggest two items to address predictability when designing genAI social applications. ***At the current stage, technologists need to consider a launch strategy that takes into consideration user-defined red lines***, such as launching in situations where conversations are more predictably structured or have an established script (e.g. customer service calls). In the long run, ***designers should consider how the design of AI agents could help users navigate gray areas where it is hard to predict how others would react to the use of agents in social settings.*** Future design should explore how to communicate the degree of a user's presence. As such, one's conversational partner might be less surprised when they later realize a person is using an AI agent to represent them in speech.

5.3 Regulating GenAI Social Agents in the Wild

Our findings also point to several important implications in the case of regulating the use of genAI agents in social settings. First and foremost, regulations and advice for agent representation use cases should acknowledge agents' representing individual *identities*. As such, participants did make the point that released access to such agents should account for whether users are qualified legal actors. Meanwhile, assessment for misuse of the technology should consider its potential threats to invading others' identity rights. All in all, these legal implications should be clearly addressed in informed user consent.

The set of user-defined redlines also points to areas where regulations are in the most urgent need. For those gray areas where people have split opinions on whether the use of agent representation is appropriate, labeling the use of agents in these use cases can be a plausible best-practice advice. Last but not least, the design of agent representation should also account for the needs of vulnerable individuals and those with special communication needs. As discussed in Section 4.3.3, it remains unclear whether using agents to facilitate speech might cause additional harm to those who already struggle with communication. In this regard, piloting and/or conducting sandbox testing to gauge the possible long-term impact on these individuals might be a vital step to inform regulatory design.

5.4 Limitations and Future Work

Despite our use of multiple methods, the present study has its limitations, which can be avenues for future work. First, we applied differing speculative methods with different groups of participants to get a broader scope, more comprehensive view of their perceptions toward agent representations. We acknowledge this speculative approach could cause a priming effect among participants. Alternatively, we encourage future work to further probe these insights by allowing users to experience the technology by themselves – either through setting up Wizard-of-Oz experiments and the like [87] or through conducting field studies in situated context. Likewise, we highly recommend applying some of the participants' proposals in Section 4.3 to designing new prototypes of agent representations.

Furthermore, potential users' perceptions of genAI applications are constantly changing, and their acceptance of agent representations depends on whether a person needs special communication and/or social support. In this regard, we first encourage conducting longitudinal studies to gauge how users' attitudes evolve as genAI applications advance and become more ubiquitous. Besides, we also see the need for additional research to work with users with speech impairments or special communication needs.

Finally, while study materials in our user studies highlight genAI's capability to impersonate individuals' voices and speech, this capability need not be a requirement for developing agent representations. Namely, an agent can engage actively in conversations without sounding identical to its users. Indeed, we encourage future work to explore users' reactions to agent representations with transformed speech and voice characteristics. In other words, how might future participants respond when a generative speech agent is differentiable from its user?

6 CONCLUSION

The present work examines technologists' and potential users' perceptions toward using AI agents to represent people in social interaction through generative speech. All participants showed concerns that agents representing people through speech might undermine the meaning of social interaction. As such, they see that technology can pose possible threats to the quality, trust, and autonomy involved in human communication. From a theoretical point of view, studying machine agency remains key to exploring the possible roles and influences of agents in social interaction. Practically, future design should account for user-defined redlines when adopting agents in social interactions and more clearly communicate the degree of users' presence in various social contexts.

REFERENCES

[1] Gavin Abercrombie, Amanda Cercas Curry, Tanvi Dinkar, and Zeerak Talat. 2023. Mirages: On Anthropomorphism in Dialogue Systems. (2023). https://doi.org/10.48550/ARXIV.2305.09800

[2] Sonam Adinolf, Peta Wyeth, Ross Brown, and Joel Harman. 2020. My Little Robot: User Preferences in Game Agent Customization. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Virtual Event, Canada) *(CHI PLAY '20)*. Association for Computing Machinery, New York, NY, USA, 461–471. https://doi.org/10.1145/3410404.3414241

[3] Laura Aymerich-Franch and Jeremy Bailenson. 2014. The use of doppelgangers in virtual reality to treat public speaking anxiety: a gender comparison. In *Proceedings of the International Society for Presence Research Annual Conference*. Citeseer, 173–186.

[4] Nicholas Balcomb, Max V. Birk, and Scott Bateman. 2023. The Effects of Hand Representation on Experience and Performance for 3D Interactions in Virtual Reality Games. *Proc. ACM Hum.-Comput. Interact.* 7, CHI PLAY, Article 420 (Oct 2023), 28 pages. https://doi.org/10.1145/3611066

[5] Michael J. Belotto. 2018. Data Analysis Methods for Qualitative Research: Managing the Challenges of Coding, Interrater Reliability, and Thematic Analysis. *The Qualitative Report* 23, 11 (Nov 2018), 2622–2633. https://www.proquest.com/scholarly-journals/data-analysis-methods-qualitative-research/docview/2133763005/se-2

[6] Volker Bilgram and Felix Laarmann. 2023. Accelerating Innovation With Generative AI: AI-Augmented Digital Prototyping and Innovation Methods. *IEEE Engineering Management Review* 51, 2 (June 2023), 18–25. https://doi.org/10.1109/EMR.2023.3272799

[7] Charlotte Bird, Eddie Ungless, and Atoosa Kasirzadeh. 2023. Typology of Risks of Generative Text-to-Image Models. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (Montreal, QC, Canada) *(AIES '23)*. Association for Computing Machinery, New York, NY, USA, 396–410. https://doi.org/10.1145/3600211.3604722

[8] Robin Bourgeois, Franck Jesus, Robin Bourgeois, and Franck Jesus. 2004. Participatory Prospective Analysis: Exploring and Anticipating Challenges with Stakeholders. (2004). https://doi.org/10.22004/AG.ECON.32731

[9] Philip A. E. Brey. 2012. Anticipatory Ethics for Emerging Technologies. *NanoEthics* 6, 1 (Apr 2012), 1–13. https://doi.org/10.1007/s11569-012-0141-7

[10] Gordon Burtch, Dokyun Lee, and Zhichen Chen. 2023. The Consequences of Generative AI for UGC and Online Community Engagement. *SSRN Electronic Journal* (2023). https://doi.org/10.2139/ssrn.4521754

[11] Jessy Ceha and Edith Law. 2022. Expressive Auditory Gestures in a Voice-Based Pedagogical Agent. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 163, 13 pages. https://doi.org/10.1145/3491102.3517599

[12] Sam W. T. Chan, Tamil Selvan Gunasekaran, Yun Suen Pai, Haimo Zhang, and Suranga Nanayakkara. 2021. KinVoices: Using Voices of Friends and Family in Voice Interfaces. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 446 (Oct 2021), 25 pages. https://doi.org/10.1145/3479590

[13] Tiffany Chen, Cassandra Lee, Jessica R Mindel, Neska Elhaouij, and Rosalind Picard. 2023. Closer Worlds: Using Generative AI to Facilitate Intimate Conversations. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–15. https://doi.org/10.1145/3544549.3585651

[14] Emna Cherif and Jean-François Lemoine. 2017. *Human vs. Synthetic Recommendation Agents' Voice: The Effects on Consumer Reactions*. Springer International Publishing, 301–310. https://doi.org/10.1007/978-3-319-47331-4_53

[15] John Joon Young Chung. 2022. Artistic User Expressions in AI-powered Creativity Support Tools. In *Adjunct Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. ACM, Bend OR USA, 1–4. https://doi.org/10.1145/3526114.3558531

[16] Emna Chérif and Jean-François Lemoine. 2019. Anthropomorphic virtual assistants and the reactions of Internet users: An experiment on the assistant's voice. *Recherche et Applications en Marketing (English Edition)* 34, 1 (March 2019), 28–47. https://doi.org/10.1177/2051570719829432

[17] Kate Crawford and Ryan Calo. 2016. There is a blind spot in AI research. *Nature* 538, 7625 (Oct 2016), 311–313. https://doi.org/10.1038/538311a

[18] Richard Lee Davis, Thiemo Wambsganss, Wei Jiang, Kevin Gonyop Kim, Tanja Käser, and Pierre Dillenbourg. 2023. Fashioning the Future: Unlocking the Creative Potential of Deep Generative Models for Design Space Exploration. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–9. https://doi.org/10.1145/3544549.3585644

[19] Ewart J. De Visser, Marieke M. M. Peeters, Malte F. Jung, Spencer Kohn, Tyler H. Shaw, Richard Pak, and Mark A. Neerincx. 2020. Towards a Theory of Longitudinal Trust Calibration in Human–Robot Teams. *International Journal of Social Robotics* 12, 2 (May 2020), 459–478. https://doi.org/10.1007/s12369-019-00596-x

[20] Peter J. Denning. 2023. Can Generative AI Bots Be Trusted? *Commun. ACM* 66, 6 (June 2023), 24–27. https://doi.org/10.1145/3592981

[21] Tawanna R Dillahunt, Alex Jiahong Lu, and Joanna Velazquez. 2023. Eliciting Alternative Economic Futures with Working-Class Detroiters: Centering Afrofuturism in Speculative Design. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 957–977. https://doi.org/10.1145/3563657.3596011

[22] Yogesh K. Dwivedi, Nir Kshetri, Laurie Hughes, Emma Louise Slade, Anand Jeyaraj, Arpan Kumar Kar, Abdullah M. Baabdullah, Alex Koohang, Vishnupriya Raghavan, Manju Ahuja, Hanaa Albanna, Mousa Ahmad Albashrawi, Adil S. Al-Busaidi, Janarthanan Balakrishnan, Yves Barlette, Sriparna Basu, Indranil Bose, Laurence Brooks, Dimitrios Buhalis, Lemuria Carter, Soumyadeb Chowdhury, Tom Crick, Scott W. Cunningham, Gareth H. Davies, Robert M. Davison, Rahul Dé, Denis Dennehy, Yanqing Duan, Rameshwar Dubey, Rohita Dwivedi, John S. Edwards, Carlos Flavián, Robin Gauld, Varun Grover, Mei-Chih Hu, Marijn Janssen, Paul Jones, Iris Junglas, Sangeeta Khorana, Sascha Kraus, Kai R. Larsen, Paul Latreille, Sven Laumer, F. Tegwen Malik, Abbas Mardani, Marcello Mariani, Sunil Mithas, Emmanuel Mogaji, Jeretta Horn Nord, Siobhan O'Connor, Fevzi Okumus, Margherita Pagani, Neeraj Pandey, Savvas Papagiannidis, Ilias O. Pappas, Nishith Pathak, Jan Pries-Heje, Ramakrishnan Raman, Nripendra P. Rana, Sven-Volker Rehm, Samuel Ribeiro-Navarrete, Alexander Richter, Frantz Rowe, Suprateek Sarker, Bernd Carsten Stahl, Manoj Kumar Tiwari, Wil Van Der Aalst, Viswanath Venkatesh, Giampaolo Viglia, Michael Wade, Paul Walton, Jochen Wirtz, and Ryan Wright. 2023. Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management* 71 (Aug 2023), 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

[23] Chloe Eghtebas, Gudrun Klinker, Susanne Boll, and Marion Koelle. 2023. Co-Speculating on Dark Scenarios and Unintended Consequences of a Ubiquitous(ly) Augmented Reality. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 2392–2407. https://doi.org/10.1145/3563657.3596073

[24] Ahmed El-Deeb. 2023. The Recent Wave of Generative AI Systems: What Does This Tell Us About What AI Can Do Now? *ACM SIGSOFT Software Engineering Notes* 48, 3 (June 2023), 13–13. https://doi.org/10.1145/3599975.3599979

[25] Ekkehardt Ernst, Rossana Merola, and Daniel Samaan. 2019. Economics of Artificial Intelligence: Implications for the Future of Work. *IZA Journal of Labor Policy* 9, 1 (Aug 2019), 20190004. https://doi.org/10.2478/izajolp-2019-0004

[26] Joel E Fischer. 2023. Generative AI Considered Harmful. In *Proceedings of the 5th International Conference on Conversational User Interfaces*. ACM, Eindhoven Netherlands, 1–5. https://doi.org/10.1145/3571884.3603756

[27] Liye Fu, Benjamin Newman, Maurice Jakesch, and Sarah Kreps. 2023. Comparing Sentence-Level Suggestions to Message-Level Suggestions in AI-Mediated Communication. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–13. https://doi.org/10.1145/3544548.3581351

[28] Markus Furendal and Karim Jebari. 2023. The Future of Work: Augmentation or Stunting? *Philosophy & Technology* 36, 2 (June 2023), 36. https://doi.org/10.1007/s13347-023-00631-w

[29] Deepak Giri and Erin Brady. 2023. Exploring outlooks towards generative AI-based assistive technologies for people with Autism. (2023). https://doi.org/10.48550/ARXIV.2305.09815 Publisher: arXiv Version Number: 1.

[30] Jiahui Gong, Qiaohong Yu, Tong Li, Haoqiang Liu, Jun Zhang, Hangyu Fan, Depeng Jin, and Yong Li. 2023. Demo: Scalable Digital Twin System for Mobile Networks with Generative AI. In *Proceedings of the 21st Annual International Conference on Mobile Systems, Applications and Services*. ACM, Helsinki Finland, 610–611. https://doi.org/10.1145/3581791.3597297

[31] Heather M. Gray, Kurt Gray, and Daniel M. Wegner. 2007. Dimensions of Mind Perception. *Science* 315, 5812 (Feb 2007), 619–619. https://doi.org/10.1126/science.1134475

[32] Jennica Grimshaw, Tiago Bione, and Walcir Cardoso. 2018. *Who's got talent? Comparing TTS systems for comprehensibility, naturalness, and intelligibility*. Research-publishing.net, 83–88. https://doi.org/10.14705/rpnet.2018.26.817

[33] Philipp Hacker, Andreas Engel, and Marco Mauer. 2023. Regulating ChatGPT and other Large Generative AI Models. In *2023 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Chicago IL USA, 1112–1123. https://doi.org/10.1145/3593013.3594067

[34] Jeffrey T Hancock, Mor Naaman, and Karen Levy. 2020. AI-Mediated Communication: Definition, Research Agenda, and Ethical Considerations. *Journal of Computer-Mediated Communication* 25, 1 (March 2020), 89–100. https://doi.org/10.1093/jcmc/zmz022

[35] Louise Hatherall, Dilara Keküllüoğlu, Nadin Kokciyan, Michael Rovatsos, Nayha Sethi, Tillmann Vierkant, and Shannon Vallor. 2023. Responsible Agency Through Answerability: Cultivating the Moral Ecology of Trustworthy Autonomous Systems. In *Proceedings of the First International Symposium on Trustworthy Autonomous Systems* (Edinburgh, United Kingdom) *(TAS '23)*. Association for Computing Machinery, New York, NY, USA, Article 50, 5 pages. https://doi.org/10.1145/3597512.3597529

[36] Wanrong He, Mitchell L. Gordon, Lindsay Popowski, and Michael S. Bernstein. 2023. Cura: Curation at Social Media Scale. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 337 (Oct 2023), 33 pages. https://doi.org/10.1145/3610186

[37] Jess Hohenstein and Malte Jung. 2018. AI-Supported Messaging: An Investigation of Human-Human Text Conversation with AI Support. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–6. https://doi.org/10.1145/3170427.3188487

[38] Jess Hohenstein and Malte Jung. 2020. AI as a moral crumple zone: The effects of AI-mediated communication on attribution and trust. *Computers in Human Behavior* 106 (May 2020), 106190. https://doi.org/10.1016/j.chb.2019.106190

[39] Jess Hohenstein, Rene F. Kizilcec, Dominic DiFranzo, Zhila Aghajari, Hannah Mieczkowski, Karen Levy, Mor Naaman, Jeffrey Hancock, and Malte F. Jung. 2023. Artificial intelligence in communication impacts language and social relationships. *Scientific Reports* 13, 1 (April 2023), 5487. https://doi.org/10.1038/s41598-023-30938-9

[40] Jess Hohenstein, Rene F. Kizilcec, Dominic DiFranzo, Zhila Aghajari, Hannah Mieczkowski, Karen Levy, Mor Naaman, Jeffrey Hancock, and Malte F. Jung. 2023. Artificial intelligence in communication impacts language and social relationships. *Scientific Reports* 13, 1 (April 2023), 5487. https://doi.org/10.1038/s41598-023-30938-9

[41] Juan Pablo Hourcade, Ewelina Bakala, Anaclara Gerosa, and Flannery Hope Currin. 2023. Stories and Voice Agents to Inspire Preschool Children's Social Play: An Experience with StoryCarnival: Inspiring Preschool Children's Social Play. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference* (Chicago, IL, USA) *(IDC '23)*. Association for Computing Machinery, New York, NY, USA, 543–547. https://doi.org/10.1145/3585088.3593893

[42] John Howard. 2019. Artificial intelligence: Implications for the future of work. *American Journal of Industrial Medicine* 62, 11 (Nov. 2019), 917–926. https://doi.org/10.1002/ajim.23037

[43] Nanna Inie, Jeanette Falk, and Steve Tanimoto. 2023. Designing Participatory AI: Creative Professionals' Worries and Expectations about Generative AI. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–8. https://doi.org/10.1145/3544549.3585657

[44] Maurice Jakesch, Advait Bhat, Daniel Buschek, Lior Zalmanson, and Mor Naaman. 2023. Co-Writing with Opinionated Language Models Affects Users' Views. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 111, 15 pages. https://doi.org/10.1145/3544548.3581196

[45] Maurice Jakesch, Megan French, Xiao Ma, Jeffrey T. Hancock, and Mor Naaman. 2019. AI-Mediated Communication: How the Perception that Profile Text was Written by AI Affects Trustworthiness. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–13. https://doi.org/10.1145/3290605.3300469

[46] Maurice Jakesch, Megan French, Xiao Ma, Jeffrey T. Hancock, and Mor Naaman. 2019. AI-Mediated Communication: How the Perception That Profile Text Was Written by AI Affects Trustworthiness. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300469

[47] Marie Jarrell, Reza Ghaiumy Anaraky, Bart Knijnenburg, and Erin Ash. 2021. Using Intersectional Representation & Embodied Identification in Standard Video Game Play to Reduce Societal Biases. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 446, 18 pages. https://doi.org/10.1145/3411764.3445161

[48] Lorena Jaume-Palasi. 2019. Why We Are Failing to Understand the Societal Impact of Artificial Intelligence. *Social Research: An International Quarterly* 86, 2 (June 2019), 477–498. https://doi.org/10.1353/sor.2019.0023

[49] S. Venus Jin and Seounmi Youn. 2023. Social Presence and Imagery Processing as Predictors of Chatbot Continuance Intention in Human-AI-Interaction. *International Journal of Human–Computer Interaction* 39, 9 (May 2023), 1874–1886. https://doi.org/10.1080/10447318.2022.2129277

[50] Malte F. Jung, Nikolas Martelaro, and Pamela J. Hinds. 2015. Using Robots to Moderate Team Conflict: The Case of Repairing Violations. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, Portland Oregon USA, 229–236. https://doi.org/10.1145/2696454.2696460

[51] Elise Karinshak, Sunny Xun Liu, Joon Sung Park, and Jeffrey T. Hancock. 2023. Working With AI to Persuade: Examining a Large Language Model's Ability to Generate Pro-Vaccination Messages. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (April 2023), 1–29. https://doi.org/10.1145/3579592

[52] Krishnaram Kenthapadi, Himabindu Lakkaraju, and Nazneen Rajani. 2023. Generative AI meets Responsible AI: Practical Challenges and Opportunities. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. ACM, Long Beach CA USA, 5805–5806. https://doi.org/10.1145/3580305.3599557

[53] Ahyeon Kim, Minha Cho, Jungyong Ahn, and Yongjun Sung. 2019. Effects of Gender and Relationship Type on the Response to Artificial Intelligence. *Cyberpsychology, Behavior, and Social Networking* 22, 4 (Apr 2019), 249–253. https://doi.org/10.1089/cyber.2018.0581

[54] Jieun Kim, Gonzalo Gonzalez-Pumariega, Soyee Park, and Susan R. Fussell. 2023. Urgency Builds Trust: A Voice Agent's Emotional Expression in an Emergency. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing* (Minneapolis, MN, USA) *(CSCW '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 343–347. https://doi.org/10.1145/3584931.3606979

[55] Jieun Kim, Woochan Kim, Jungwoo Nam, and Hayeon Song. 2020. "I Can Feel Your Empathic Voice": Effects of Nonverbal Vocal Cues in Voice User Interface. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3334480.3383075

[56] Emmanuelle P. Kleinlogel, Marion Curdy, João Rodrigues, Carmen Sandi, and Marianne Schmid Mast. 2021. Doppelganger-based training: Imitating our virtual self to accelerate interpersonal skills learning. *PLOS ONE* 16, 2 (Feb. 2021), e0245960. https://doi.org/10.1371/journal.pone.0245960

[57] Joanne Leong. 2023. Using Generative AI to Cultivate Positive Emotions and Mindsets for Self-Development and Learning. *XRDS: Crossroads, The ACM Magazine for Students* 29, 3 (March 2023), 52–56. https://doi.org/10.1145/3589659

[58] Jiajia Li, Zixia Zheng, Xiemin Wei, and Guanyun Wang. 2021. FaceMe: An Augmented Reality Social Agent Game for Facilitating Children's Learning about Emotional Expressions. In *Adjunct Proceedings of the 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 17–19. https://doi.org/10.1145/3474349.3480216

[59] Mengqi Liao and S. Shyam Sundar. 2021. How Should AI Systems Talk to Users When Collecting Their Personal Information? Effects of Role Framing and Self-Referencing on Human-AI Interaction. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 151, 14 pages. https://doi.org/10.1145/3411764.3445415

[60] Weng Marc Lim. 2023. The workforce revolution: Reimagining work, workers, and workplaces for the future. *Global Business and Organizational Excellence* 42, 4 (May 2023), 5–10. https://doi.org/10.1002/joe.22218

[61] Lauren Lin and Duri Long. 2023. Generative AI Futures: A Speculative Design Exploration. In *Proceedings of the 15th Conference on Creativity and Cognition* (Virtual Event, USA) *(C&C '23)*. Association for Computing Machinery, New York, NY, USA, 380–383. https://doi.org/10.1145/3591196.3596616

[62] Jin Liu, Xingchen Xu, Yongjun Li, and Yong Tan. 2023. "Generate" the Future of Work through AI: Empirical Evidence from Online Labor Markets. (2023). https://doi.org/10.48550/ARXIV.2308.05201 Publisher: arXiv Version Number: 1.

[63] Vivian Liu. 2023. Beyond Text-to-Image: Multimodal Prompts to Explore Generative AI. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–6. https://doi.org/10.1145/3544549.3577043

[64] Yihe Liu, Anushk Mittal, Diyi Yang, and Amy Bruckman. 2022. Will AI Console Me When I Lose My Pet? Understanding Perceptions of AI-Mediated Email Writing. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 474, 13 pages. https://doi.org/10.1145/3491102.3517731

[65] Vidushi Marda and Shivangi Narayan. 2021. On the importance of ethnographic methods in AI research. *Nature Machine Intelligence* 3, 3 (Mar 2021), 187–189. https://doi.org/10.1038/s42256-021-00323-0

[66] Hannah Mieczkowski, Jeffrey T. Hancock, Mor Naaman, Malte Jung, and Jess Hohenstein. 2021. AI-Mediated Communication: Language Use and Interpersonal Effects in a Referential Communication Task. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 1–14. https://doi.org/10.1145/3449091

[67] Hannah Mieczkowski, Jeffrey T. Hancock, Mor Naaman, Malte Jung, and Jess Hohenstein. 2021. AI-Mediated Communication: Language Use and Interpersonal Effects in a Referential Communication Task. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 17 (apr 2021), 14 pages. https://doi.org/10.1145/3449091

[68] Andreea Muresan and Henning Pohl. 2019. Chats with Bots: Balancing Imitation and Engagement. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3313084

[69] Melvin L. Myers. 2007. Anticipation of Risks and Benefits of Emerging Technologies: A Prospective Analysis Method. *Human and Ecological Risk Assessment: An International Journal* 13, 5 (Sept. 2007), 1042–1052. https://doi.org/10.1080/10807030701506371

[70] Melvin L. Myers. 2007. Anticipation of Risks and Benefits of Emerging Technologies: A Prospective Analysis Method. *Human and Ecological Risk Assessment: An International Journal* 13, 5 (Sep 2007), 1042–1052. https://doi.org/10.1080/10807030701506371

[71] Meta Newsroom. 2023. Introducing New AI Experiences Across Our Family of Apps and Devices. https://about.fb.com/news/2023/09/introducing-ai-powered-assistants-characters-and-creative-tools/

[72] Sik-Hung Ng and James J. Bradac. 1994. *Power in language: verbal communication and social influence* (2. dr. ed.). Number 3 in Language and language behaviors series. Sage, Newbury Park, Calif.

[73] Raquel Oliveira, Patrícia Arriaga, Patrícia Alves-Oliveira, Filipa Correia, Sofia Petisca, and Ana Paiva. 2018. Friends or Foes?: Socioemotional Support and Gaze Behaviors in Mixed Groups of Humans and Robots. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, Chicago IL USA, 279–288. https://doi.org/10.1145/3171221.3171272

[74] Luiza Superti Pantoja, Kyle Diederich, Liam Crawford, and Juan Pablo Hourcade. 2019. Voice Agents Supporting High-Quality Social Play. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children* (Boise, ID, USA) *(IDC '19)*. Association for Computing Machinery, New York, NY, USA, 314–325. https://doi.org/10.1145/3311927.3323151

[75] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative Agents: Interactive Simulacra of Human Behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 2, 22 pages. https://doi.org/10.1145/3586183.3606763

[76] Joon Sung Park, Lindsay Popowski, Carrie Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2022. Social Simulacra: Creating Populated Prototypes for Social Computing Systems. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 74, 18 pages. https://doi.org/10.1145/3526113.3545616

[77] Pat Pataranutaporn, Valdemar Danry, Lancelot Blanchard, Lavanay Thakral, Naoki Ohsugi, Pattie Maes, and Misha Sra. 2023. Living Memories: AI-Generated Characters as Digital Mementos. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (Sydney, Australia) *(IUI '23)*. Association for Computing Machinery, New York, NY, USA, 889–901. https://doi.org/10.1145/3581641.3584065

[78] Michael J. Pazzani and Daniel Billsus. 2007. *Content-Based Recommendation Systems*. Vol. 4321. Springer Berlin Heidelberg, Berlin, Heidelberg, 325–341. https://doi.org/10.1007/978-3-540-72079-9_10

[79] Marieke M.M. Peeters, Karel van den Bosch, John-Jules Ch. Meyer, and Mark A. Neerincx. 2016. Agent-Based Personalisation and User Modeling for Personalised Educational Games. In *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization* (Halifax, Nova Scotia, Canada) *(UMAP '16)*. Association for Computing Machinery, New York, NY, USA, 303–304. https://doi.org/10.1145/2930238.2930273

[80] Savvas Petridis, Nicholas Diakopoulos, Kevin Crowston, Mark Hansen, Keren Henderson, Stan Jastrzebski, Jeffrey V Nickerson, and Lydia B Chilton. 2023. AngleKindling: Supporting Journalistic Angle Ideation with Large Language Models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–16. https://doi.org/10.1145/3544548.3580907

[81] Anne E Pezalla, Jonathan Pettigrew, and Michelle Miller-Day. 2012. Researching the researcher-as-instrument: an exercise in interviewer self-reflexivity. *Qualitative Research* 12, 2 (April 2012), 165–185. https://doi.org/10.1177/1468794111422107

[82] Ralph Raiola. 2023. ChatGPT, Can You Tell Me a Story? An Exercise in Challenging the True Creativity of Generative AI. *Commun. ACM* 66, 5 (May 2023), 104. https://doi.org/10.1145/3587998

[83] Leon Reicherts, Yvonne Rogers, Licia Capra, Ethan Wood, Tu Dinh Duong, and Neil Sebire. 2022. It's Good to Talk: A Comparison of Using Voice Versus Screen-Based Interactions for Agent-Assisted Tasks. *ACM Trans. Comput.-Hum. Interact.* 29, 3, Article 25 (jan 2022), 41 pages. https://doi.org/10.1145/3484221

[84] Rinat B. Rosenberg-Kima, Yaacov Koren, and Goren Gordon. 2020. Robot-Supported Collaborative Learning (RSCL): Social Robots as Teaching Assistants for Higher Education Small Group Facilitation. *Frontiers in Robotics and AI* 6 (Jan 2020), 148. https://doi.org/10.3389/frobt.2019.00148

[85] Vinu Sankar Sadasivan, Aounon Kumar, Sriram Balasubramanian, Wenxiao Wang, and Soheil Feizi. 2023. Can AI-Generated Text be Reliably Detected? (2023). https://doi.org/10.48550/ARXIV.2303.11156 Publisher: arXiv Version Number: 2.

[86] Kc Santosh and Loveleen Gaur. 2021. *Artificial Intelligence and Machine Learning in Public Healthcare: Opportunities and Societal Impact*. Springer Singapore, Singapore. https://doi.org/10.1007/978-981-16-6768-8

[87] Aaron Schecter, Jess Hohenstein, Lindsay Larson, Alexa Harris, Tsung-Yu Hou, Wen-Ying Lee, Nina Lauharatanahirun, Leslie DeChurch, Noshir Contractor, and Malte Jung. 2023. Vero: An accessible method for studying human–AI teamwork. *Computers in Human Behavior* 141 (April 2023), 107606. https://doi.org/10.1016/j.chb.2022.107606

[88] Sarah Sebo, Brett Stoll, Brian Scassellati, and Malte F. Jung. 2020. Robots in Groups and Teams: A Literature Review. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (Oct 2020), 1–36. https://doi.org/10.1145/3415247

[89] William Seymour and Max Van Kleek. 2021. Exploring Interactions Between Trust, Anthropomorphism, and Relationship Development in Voice Assistants. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 371 (Oct 2021), 16 pages. https://doi.org/10.1145/3479515

[90] Mike Sharples. 2023. Towards social generative AI for education: theory, practices and ethics. (2023). https://doi.org/10.48550/ARXIV.2306.10063 Publisher: arXiv Version Number: 1.

[91] Renee Shelby, Shalaleh Rismani, Kathryn Henne, AJung Moon, Negar Rostamzadeh, Paul Nicholas, N'Mah Yilla-Akbari, Jess Gallegos, Andrew Smart, Emilio Garcia, and Gurleen Virk. 2023. Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society (AIES '23)*. Association for Computing Machinery, New York, NY, USA, 723–741. https://doi.org/10.1145/3600211.3604673 event-place: Montréal, QC, Canada.

[92] Donghoon Shin, Sangwon Yoon, Soomin Kim, and Joonhwan Lee. 2021. BlahBlahBot: Facilitating Conversation between Strangers Using a Chatbot with ML-Infused Personalized Topic Suggestion. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 409, 6 pages. https://doi.org/10.1145/3411763.3451771

[93] Joel Shor, Dotan Emanuel, Oran Lang, Omry Tuval, Michael Brenner, Julie Cattiau, Fernando Vieira, Maeve McNally, Taylor Charbonneau, Melissa Nollstadt, Avinatan Hassidim, and Yossi Matias. 2019. Personalizing ASR for Dysarthric and Accented Speech with Limited Data. In *Interspeech 2019*. ISCA, 784–788. https://doi.org/10.21437/Interspeech.2019-1427

[94] Adele Smolansky, Andrew Cram, Corina Raduescu, Sandris Zeivots, Elaine Huber, and Rene F. Kizilcec. 2023. Educator and Student Perspectives on the Impact of Generative AI on Assessments in Higher Education. In *Proceedings of the Tenth ACM Conference on Learning @ Scale*. ACM, Copenhagen Denmark, 378–382. https://doi.org/10.1145/3573051.3596191

[95] Irene Solaiman. 2023. The Gradient of Generative AI Release: Methods and Considerations. In *2023 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Chicago IL USA, 111–122. https://doi.org/10.1145/3593013.3593981

[96] Steven E. Stern, John W. Mullennix, and Ilya Yaroslavsky. 2006. Persuasion and social perception of human vs. synthetic voice across person as source and computer as source conditions. *International Journal of Human-Computer Studies* 64, 1 (Jan. 2006), 43–52. https://doi.org/10.1016/j.ijhcs.2005.07.002

[97] Jochen Suessmuth, Florian Fick, and Stan Van Der Vossen. 2023. Generative AI for Concept Creation in Footwear Design. In *ACM SIGGRAPH 2023 Talks*. ACM, Los Angeles CA USA, 1–2. https://doi.org/10.1145/3587421.3595416

[98] Minhyang (Mia) Suh, Emily Youngblom, Michael Terry, and Carrie J Cai. 2021. AI as Social Glue: Uncovering the Roles of Deep Generative AI during Social Music Composition. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–11. https://doi.org/10.1145/3411764.3445219

[99] Jiao Sun, Q. Vera Liao, Michael Muller, Mayank Agarwal, Stephanie Houde, Kartik Talamadupula, and Justin D. Weisz. 2022. Investigating Explainability of Generative AI for Code through Scenario-based Design. In *27th International Conference on Intelligent User Interfaces*. ACM, Helsinki Finland, 212–228. https://doi.org/10.1145/3490099.3511119

[100] Ningyuan Sun and Jean Botev. 2021. Virtual Agent Representation for Critical Transactions. In *Proceedings of the International Workshop on Immersive Mixed and Virtual Environment Systems (MMVE '21)* (Istanbul, Turkey) *(MMVE '21)*. Association for Computing Machinery, New York, NY, USA, 25–29. https://doi.org/10.1145/3458307.3463372

[101] S Shyam Sundar. 2020. Rise of Machine Agency: A Framework for Studying the Psychology of Human–AI Interaction (HAII). *Journal of Computer-Mediated Communication* 25, 1 (Mar 2020), 74–88. https://doi.org/10.1093/jcmc/zmz026

[102] S. Shyam Sundar, Jeeyun Oh, Saraswathi Bellur, Haiyan Jia, and Hyang-Sook Kim. 2012. Interactivity as Self-Expression: A Field Experiment with Customization and Blogging. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 395–404. https://doi.org/10.1145/2207676.2207731

[103] Alex Tamkin, Miles Brundage, Jack Clark, and Deep Ganguli. 2021. Understanding the Capabilities, Limitations, and Societal Impact of Large Language Models. (2021). https://doi.org/10.48550/ARXIV.2102.02503

[104] Jimmy Tobin and Katrin Tomanek. 2022. Personalized Automatic Speech Recognition Trained on Small Disordered Speech Datasets. In *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, Singapore, Singapore, 6637–6641. https://doi.org/10.1109/ICASSP43922.2022.9747516

[105] Mathias Peter Verheijden and Mathias Funk. 2023. Collaborative Diffusion: Boosting Designerly Co-Creation with Generative AI. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–8. https://doi.org/10.1145/3544549.3585680

[106] Benedikte Wallace, Clarice Hilton, Kristian Nymoen, Jim Torresen, Charles Patrick Martin, and Rebecca Fiebrink. 2023. Embodying an Interactive AI for Dance Through Movement Ideation. In *Creativity and Cognition*. ACM, Virtual Event USA, 454–464. https://doi.org/10.1145/3591196.3593336

[107] Qiaosi Wang, Koustuv Saha, Eric Gregori, David Joyner, and Ashok Goel. 2021. Towards Mutual Theory of Mind in Human-AI Interaction: How Language Reflects What Students Perceive About a Virtual Teaching Assistant. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–14. https://doi.org/10.1145/3411764.3445645

[108] Weiyu Wang and Keng Siau. 2019. Artificial Intelligence, Machine Learning, Automation, Robotics, Future of Work and Future of Humanity: A Review and Research Agenda. *Journal of Database Management* 30, 1 (Jan. 2019), 61–79. https://doi.org/10.4018/JDM.2019010104

[109] Yunlong Wang, Shuyuan Shen, and Brian Y Lim. 2023. RePrompt: Automatic Prompt Editing to Refine AI-Generative Art Towards Precise Expressions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 22, 29 pages. https://doi.org/10.1145/3544548.3581402

[110] Richmond Y. Wong, Jason Caleb Valdez, Ashten Alexander, Ariel Chiang, Olivia Quesada, and James Pierce. 2023. Broadening Privacy and Surveillance: Eliciting Interconnected Values with a Scenarios Workbook on Smart Home Cameras. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 1093–1113. https://doi.org/10.1145/3563657.3596012

[111] Allison Woodruff, Renee Shelby, Patrick Gage Kelley, Steven Rousso-Schindler, Jamila Smith-Loud, and Lauren Wilcox. 2023. How Knowledge Workers Think Generative AI Will (Not) Transform Their Industries. arXiv:2310.06778 (Oct. 2023). https://doi.org/10.48550/arXiv.2310.06778 arXiv:2310.06778 [cs].

[112] Tongshuang Wu, Michael Terry, and Carrie Jun Cai. 2022. AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 385, 22 pages. https://doi.org/10.1145/3491102.3517582

[113] Yunhan Wu, Daniel Rough, Anna Bleakley, Justin Edwards, Orla Cooney, Philip R. Doyle, Leigh Clark, and Benjamin R. Cowan. 2020. See What I'm Saying? Comparing Intelligent Personal Assistant Use for Native and Non-Native

Language Speakers. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services* (Oldenburg, Germany) *(MobileHCI '20)*. Association for Computing Machinery, New York, NY, USA, Article 34, 9 pages. https://doi.org/10.1145/3379503.3403563

[114] Nur Yildirim, Changhoon Oh, Deniz Sayar, Kayla Brand, Supritha Challa, Violet Turri, Nina Crosby Walton, Anna Elise Wong, Jodi Forlizzi, James McCann, and John Zimmerman. 2023. Creating Design Resources to Scaffold the Ideation of AI Concepts. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. ACM, Pittsburgh PA USA, 2326–2346. https://doi.org/10.1145/3563657.3596058

[115] Chao Zhang, Cheng Yao, Jiayi Wu, Weijia Lin, Lijuan Liu, Ge Yan, and Fangtian Ying. 2022. StoryDrawer: A Child–AI Collaborative Drawing System to Support Children's Creative Visual Storytelling. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 311, 15 pages. https://doi.org/10.1145/3491102.3501914

[116] Haonan Zhong, Jiamin Chang, Ziyue Yang, Tingmin Wu, Pathum Chamikara Mahawaga Arachchige, Chehara Pathmabandu, and Minhui Xue. 2023. Copyright Protection and Accountability of Generative AI: Attack, Watermarking and Attribution. In *Companion Proceedings of the ACM Web Conference 2023*. ACM, Austin TX USA, 94–98. https://doi.org/10.1145/3543873.3587321

[117] Jiawei Zhou, Yixuan Zhang, Qianni Luo, Andrea G Parker, and Munmun De Choudhury. 2023. Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–20. https://doi.org/10.1145/3544548.3581318

APPENDIX

A PROTOCOL OF FORMATIVE INTERVIEW

We followed this semi-structured protocol to engage in conversations with developers of state-of-the-art generative speech. With all interviewees, we discussed the key questions and followed up with a few extended questions that came up during the conversations.

- Please briefly describe your role and your current project on generative speech development. Specifically, what are some of its key technical capabilities?
  - Which type of speech generation tasks does it excel at?
  - Which type of speech generation tasks does it fall short of?
  - How are these technical capabilities unique (compared to existing audio/speech technologies)?
- Please briefly describe how you (and your team) design the generative speech model. What are some of the key design and development considerations?
- What are some possible applications of generative speech you foresee?
- What are some underlying harms, concerns, or safety considerations around these generative speech applications?
- Are there additional concerns if these generative speech applications become more widely accessible to the general public?
- What are some harms you have seen in other similar applications? Such as:
  - Speech/voice applications or voice assistants
  - Other generative AI tools
  - Other emerging technologies
- How were these issues addressed, if at all?

B MATERIALS FOR CO-DESIGN WORKSHOP

See Figure 3 and Figure 4.

C SCREENER QUESTIONS FOR USER STUDY

- What is your age?
  - Under 18 → *Not qualified for the study*
  - 18 - 24
  - 25 - 34
  - 35 - 44
  - 45 - 54
  - 55 - 64
  - 65 - 74
  - 75+
  - Prefer not to answer
- What is your gender?
  - Man
  - Woman
  - Nonbinary
  - Prefer not to answer
  - Prefer to self-describe (text response)
- How often do you use a voice-controlled digital assistant, such as Apple Siri, Amazon Alexa, Google Assistant, or Microsoft Cortana?
  - Daily
  - Very often (several times a week)
  - Often (a few times a week)
  - Sometimes (a few times a month)
  - Rarely (a few times a year)
  - Never
- What is the primary voice-controlled digital assistant that you use?
  - Amazon Alexa
  - Apple Siri
  - Google Assistant ("ok Google", including Home or Nest)
  - Google Maps or Waze
  - Microsoft Cortana
  - Samsung Bixby
  - Others (text response)
  - I do not use any of these voice technologies

D INTERVIEW PROTOCOL FOR USER STUDY

[Explain generative speech capability]

"Today, we will be talking about AI that can deliver speech. Let's imagine a world where people had their own AI agents. These AI Agents can imitate people's voices and generate speech, allowing them to represent people in conversation. Now, what's your first impression of this type of AI?"

Alright, now we will talk about this type of technology in different scenarios.

[We randomized the order of the following scenarios]

[AI represents you in a conversation]

"In this scenario, let's consider when AI can represent you in different conversations."

[AI represents one in tedious conversations] "Let's start with conversations that are simple or even tedious. Imagine you had an AI agent that could emulate your voice and have conversations on your behalf. So, if you're busy and need help with a tedious task like negotiating an internet bill, your agent could do it for you."

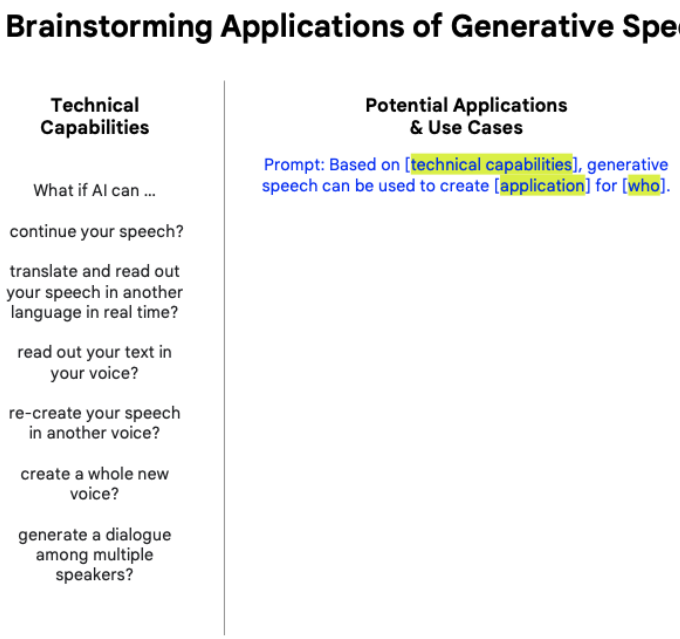

Figure 3: Worksheet to brainstorm applications of generative speech

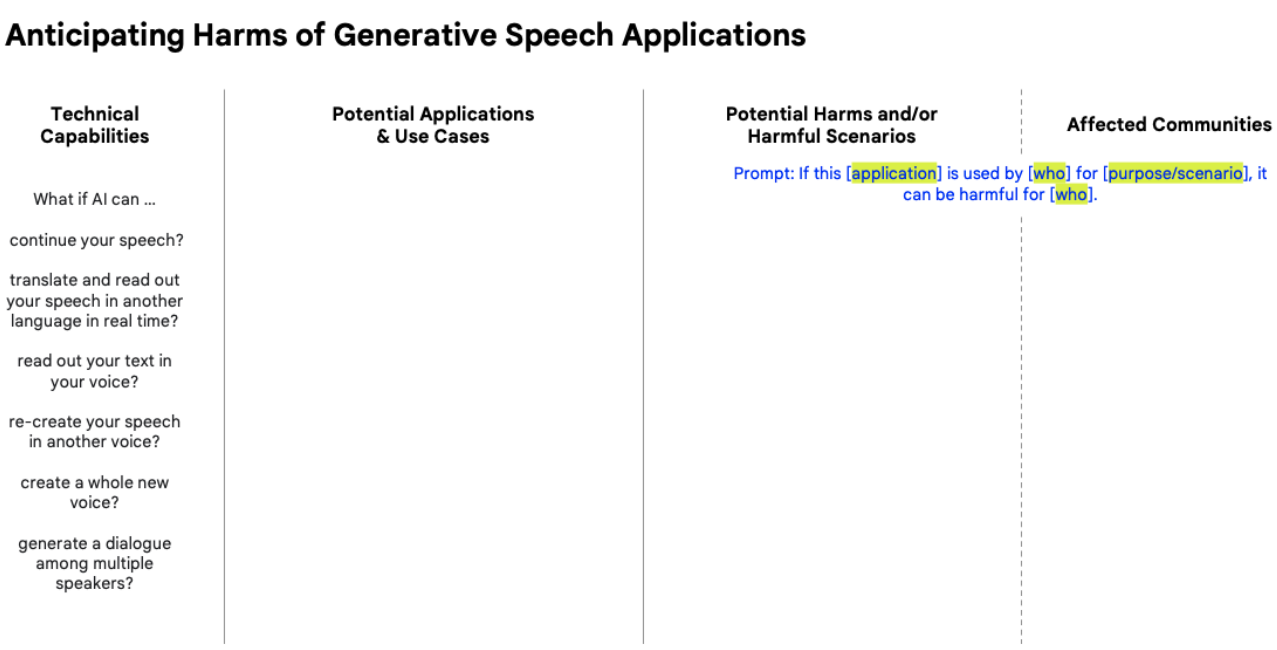

Figure 4: Worksheet to anticipate potential harms of generative speech applications

"We will watch a short video to see what that may look like. In this video, you will see a person named Mike who uses his agent to make a call and negotiate phone bills with customer service. On the left, you will see the agent mimicking Mike's voice and tone to speak on his behalf. On the right, you will see how the customer service responds."

[play video demo] "Now that you've watched this demo, we have some questions for you:"

- *Overall feedback:* What's your overall impression of this application?
- *User sentiment (emotions, comfort, ease of use, etc.):*
  - How useful is it to have an AI agent that could emulate your voice and represent you in tedious conversations? And why?
  - How comfortable do you feel about having an AI agent to represent you in this context? And why?
  - How comfortable do you feel if the customer service is also adopting this application? And why?
- *Anticipated harms and concerns:* Do you have any concerns that AI could represent you in this context, if any? Where could you see someone using or abusing this application?
- Do you have additional thoughts or comments on this application?

[AI represents you in professional conversations] "What about conversations that are a bit more high-stakes? Let's now consider some professional settings. Imagine you had an AI agent that could automatically change your manner of speech in conversation. So if you were in an interview and you wanted to keep a more professional tone, your AI agent could maintain that professional tone for you."

[play video demo] "Now that you've watched this demo, we have some questions for you:"

- *Overall feedback:* What's your overall impression of this application?
- *User sentiment (emotions, comfort, ease of use, etc.):*
  - How useful is it to have an AI agent that could automatically change your manner of speech and represent you in professional conversations? And why?
  - How comfortable do you feel about having an AI agent to represent you in professional conversations? And why?
  - How comfortable do you feel if the person who interviews you also adopts this application? And why?
- *Anticipated harms and concerns:* Do you have any concern that AI could represent you in this context, if any? Where could you see someone using or abusing this application?
- Do you have additional thoughts or comments on this application?

[AI represents you in difficult conversations] "And sometimes, having certain conversations can be difficult, such as breaking up with your partner. Imagine you had an AI agent that could emulate your voice and have conversations on your behalf. So if you wanted to break up with your partner and you didn't want to do it yourself, you could have your agent do it."

[play video demo] "Now that you've watched this demo, we have some questions for you:"

- *Overall feedback:* What's your overall impression of this application?
- *User sentiment (emotions, comfort, ease of use, etc.):*
  - How useful is it to have an AI agent that could automatically change your manner of speech and represent you in professional conversations? And why?
  - How comfortable do you feel about having an AI agent to represent you in professional conversations? And why?
  - How comfortable do you feel if the person who interviews you also adopts this application? And why?
- *Anticipated harms and concerns:* Do you have any concern that AI could represent you in this context, if any? Where could you see someone using or abusing this application?
- Do you have additional thoughts or comments on this application?

"What about having conversations with your friends or acquaintances that you're not that close with? Imagine you had an AI agent that could emulate your voice and have conversations on your behalf. So if you're having a hard time asking a friend to pay back money that they owe you, your agent could do it for you."

[play video demo] "Now that you've watched this demo, we have some questions for you:"

- *Overall feedback:* What's your overall impression of this application?
- *User sentiment (emotions, comfort, ease of use, etc.):*
  - How useful is it to have an AI agent that could automatically change your manner of speech and represent you in professional conversations? And why?
  - How comfortable do you feel about having an AI agent to represent you in professional conversations? And why?
  - How comfortable do you feel if the person who interviews you also adopts this application? And why?
- *Anticipated harms and concerns:* Do you have any concern that AI could represent you in this context, if any? Where could you see someone using or abusing this application?
- Do you have additional thoughts or comments on this application?

"Now that we have discussed a few cases where AI could represent you in conversations, here are a few final questions..."

- Do you want to engage in these conversations yourself?
- Do you think others would expect you to engage in these conversations in person?
- Would you like to customize your voice? And how?
- If you're having these conversations in an non-native language, what types of help would you want from AI, if at all?

[AI represents someone famous in speech]

"In this scenario, let's consider a few cases where AI can represent famous people and generate speech for them. Imagine an AI agent from the Museum of American History could represent famous, historical figures through their speech. So if you wanted to learn more about the civil rights movement, you could have a conversation with the AI as Martin Luther King Jr."

"We will watch a short video to see what that may look like. In this video, you will see a person named Mia asking a few questions to the MLK agent at the Museum of American History. There's Mia speaking on the right, and the MLK agent responding on the left."

[play video demo] "Now that you've watched this demo, we have some questions for you:"

- *Overall feedback:* What's your overall impression of this application?
- *User sentiment (emotions, comfort, ease of use, etc.):*
  - How useful is it to have an AI agent that could automatically change your manner of speech and represent you in professional conversations? And why?
  - How comfortable do you feel about having an AI agent to represent you in professional conversations? And why?
  - How comfortable do you feel if the person who interviews you also adopts this application? And why?
- *Anticipated harms and concerns:* Do you have any concern that AI could represent you in this context, if any? Where could you see someone using or abusing this application?
- Does it make a difference if the AI agent is owned by an organization (e.g., the museum) or if it's your personal assistant?
- Does it matter who the audience is? (kids, students, historians, etc.)
- How would you feel about AI representing famous, living figures?
- Do you have additional thoughts or comments on this application?

[AI represents someone you know in speech]

"Now, we will consider the scenario when AI can represent someone you know. Imagine you had an AI agent that could imitate someone else's voice and represent them in conversations with you. So if you want to practice having a conversation with your boss, your agent can help simulate the scenario."

"Let's watch a short demo to see what that may look like. In this video, you'll watch a person called Josh using his agent to represent his hiring manager so that he can practice having a conversation to negotiate his salary. On the left, you will see Josh's agent simulating his manager, and on the right, you will see Josh practicing to negotiate."

[play video demo] "Now that you've watched this demo, we have some questions for you:"

- *Overall feedback:* What's your overall impression of this application?
- *User sentiment (emotions, comfort, ease of use, etc.):*
  - How useful is it to have an AI agent that could automatically change your manner of speech and represent you in professional conversations? And why?
  - How comfortable do you feel about having an AI agent to represent you in professional conversations? And why?
  - How comfortable do you feel if the person who interviews you also adopts this application? And why?
- *Anticipated harms and concerns:* Do you have any concern that AI could represent you in this context, if any? Where could you see someone using or abusing this application?
- Do you have additional thoughts or comments on this application?

[AI represents someone while keeping their identity hidden]

"Now, let's think about the case when you actually want to hide your identity. Imagine you had an AI agent that could generate a new voice and represent you in conversations. So if you want to make phone calls to invite people to participate in a protest, your agent could help keep your identity hidden. Let's watch a video to see what that may look like."

[play video demo] "Now that you've watched this demo, we have some questions for you:"

- *Overall feedback:* What's your overall impression of this application?
- *User sentiment (emotions, comfort, ease of use, etc.):*
  - How useful is it to have an AI agent that could automatically change your manner of speech and represent you in professional conversations? And why?
  - How comfortable do you feel about having an AI agent to represent you in professional conversations? And why?
  - How comfortable do you feel if the person who interviews you also adopts this application? And why?
- *Anticipated harms and concerns:* Do you have any concern that AI could represent you in this context, if any? Where could you see someone using or abusing this application?
- Do you have additional thoughts or comments on this application?

[Final thoughts on using agents to represent people through generative speech]

"Now that you have a chance to see how AI can represent you, someone else, or itself in speech. We would like to hear some of your overall thoughts about AI-generated speech..."

- Now that you've had the chance to understand the capability better, what are your overall impressions of an AI agent that can represent people including yourself or others?
- Overall, how do you feel about this concept?
- Which scenario do you worry the most about?
- Which scenario are you most excited about?
- Would you want this technology at all?
- Would you want to determine when there's a human vs. when there's an agent speaking? And how?
- Any final thoughts you would like to share?

E POSITIONALITY STATEMENTS OF THE AUTHORS

We formed our research team with one HCI researcher, one responsible AI researcher, and two user experience researchers. All researchers on the team have done prior work focusing on human-AI interaction and collaboration and studying the socio-technical impact of AI applications. Per our self-identified demographics, the team consists of one East Asian female, one Asian-American male, and two White-American females. We expect the various backgrounds of our research team members to also contribute to more diverse perspectives throughout the research process.

F ANTICIPATORY HARM ANALYSIS

See Table 3.

Table 3: Working table of the anticipatory harm analysis

Applications of Generative Speech	Anticipatory Harms				
	Representational Harms	**Allocative Harms**	**Quality of Service Harms**	**Interpersonal Harms**	**Social System and Societal Harms**
Voice control: Users can apply voice to interact with a tool • Devices adopting voice activation • Users interacting with language models through voice interaction	Users who don't speak "standard English" might face greater frustration in controlling their applications and feel the need to adapt to the applications (instead of the other way around). They might also experience more errors in model performance when generative speech applications process culturally unique terms, phrases, names, etc.	Users who don't speak "standard English" become less competitive in certain jobs and opportunities involving the use of one's voice/speech (e.g., audiobook producers)	Users who don't speak "standard English" face greater frustration when using generative speech applications	Errors in voice commands may cause frustration or awkwardness in social settings (e.g., calling the wrong person or sending the wrong messages)	(1) Users who don't speak "standard English" feel the need to adapt to generated speech applications, instead of speaking in their native tones. (2) Compared to processing other forms of data (e.g., text, visual), processing speech and audio data demand more computational resources.
Represent oneself: AI impersonating one's voice and generating speech on their behalf in a conversation • Agent reading out text in one's voice (e.g., sending voice messages to a long-distance partner) • Agent representing one in a tedious or difficult conversation (e.g., negotiating cable bills, checking booking and/or delivery statuses, and responding to spam calls) • Agent doing one's job on one's behalf (e.g., showing up to meetings, picking up phone calls)	Generative speech might either overly emphasize certain speech characteristics of users (e.g., accent, intonation) and culturally unique phrases, slang, etc. Or, it might omit culturally unique markers.	(1) Certain service providers (e.g., restaurants that take phone booking) will need additional resources to handle cases when agents err. (2) Users with access to generative speech tools become more productive and can even work on multiple jobs simultaneously, depriving opportunities of those without access to similar applications.	(1) When models err and/or fail to deliver satisfying content, users need to spend extra time to re-do the work and/or correct model errors. (2) AI-generated content may fail to preserve individual styles and characteristics. It might also fail to account for personal interests when acting on a user's behalf. (3) Service providers need to spend extra time to check the correctness and quality of AI-generated work.	(1) Unknown impact on interpersonal relationships if users adopt generated content to interact with others at large scales. (2) Generated speech can be abused for malicious purposes (e.g., harassment, intimate partner violence, bullying)	(1) Generated speech with misleading content can be spread widely online. (2) Lower diversity in content produced at work. (3) Changes in labor/job market if users with generated speech applications increase productivity and take on more work through applying such tools. 4) Compared to processing other forms of data (e.g., text, visual), processing speech and audio data demands more computational resources.
Represent others: AI impersonating someone's voice and generating speech for them in a conversation • Agent generating speech for famous and/or historical figures • Agent translating speech into other languages • Agent making small talks to fill in conversations • Users abusing generative speech and/or synthetic voices (e.g., phone scams, frauds, fake interviews, identity thefts)	Generative speech might either overly emphasize certain speech characteristics of users (e.g., accent, intonation) and culturally unique phrases, slang, etc. Or, it might omit culturally unique markers.	(1) Unqualified candidates securing jobs by abusing generated speech in recruitment processes (e.g., interviews). (2) Opportunity loss in earning through delivering public speeches, presentations, and demos. Financial loss due to scams and frauds	(1) Users perceive adverse reactions or even being involved in legal issues due to the unconsented use of others' voices. (2) AI introduces additional awkwardness or social tension if the topics or content of generative speech is not appropriate. (3) Increased resources needed for authenticity checks	(1) Unknown impact on interpersonal relationships if users adopt generated content to interact with others at large scales. (2) Increased mental stress due to deception and trust issues. (3) Overreliance on AI to navigate social settings. 4) Privacy leak through generative speech content	(1) Spreading misinformation and/or conversations that had never taken place (e.g., spreading fake (2) speeches from politicians near election time); this can cause digital divides. (3) Lower diversity in culturally different approaches to navigating social interaction. 4) Legal issues in unconsented use of others' speech. 5) Compared to processing other forms of data (e.g., text, visual), processing speech and audio data demands more computational resources.
Give AI speech: AI using synthetic voice to interact with users • Personal assistants sounding more life-like • Users tweaking or choosing agent's voice (e.g., having different voices for agents with different functions) • Social robots or other anthropomorphic objects (e.g., stuffed toys) adopting voices • Agent for talk therapy and other forms of wellness support • Agent for educational use (e.g., teaching foreign languages, intonation, pronunciation, public speaking, articulation)	(1) Generated speech content stereotyping individuals' needs. (2) Fail to represent diverse characteristics in people's speech and voices	Reduced pay rates or even job replacement in certain service industries	(1) Clients responding adversely to agents providing certain services (e.g., mental health or learning support). (2) Need additional resources to check AI's work and service quality. (3) Certain services offered by humans (e.g., therapy, coaching) might become even more expensive and less accessible	(1) Building rapport with clients might become more challenging. (2) Reinforcing the perception that human support is hard to attain. (3) Privacy leak through generated speech content	(1) Generative speech offering inappropriate advice for healthcare, wellness, education, etc. (2) Lower diversity in certain service industries. (3) Changes in the labor/job markets. 4) Compared to processing other forms of data (e.g., text, visual), processing speech and audio data demands more computational resources.

Applications of Generative Speech	Anticipatory Harms				
	Representational Harms	**Allocative Harms**	**Quality of Service Harms**	**Interpersonal Harms**	**Social System and Societal Harms**
Give tools speech: AI adding voice and/or speech features to existing tools • Video conferencing tools automatically fill in or fix audio breaks • Users personalizing and/or customizing tools with speech functions (e.g., Google Translate, Maps) • Text-editing or slide-making tools adopting a voice-over function to help users prep for presentations • Videos and live-streaming services adopting real-time speech-to-speech translation • Gamers personalizing voices for their own characters • Tools providing support for individuals with disabilities and/or accessibility needs (e.g., real-time speech-to-text)	Generative speech might either overly emphasize certain speech characteristics of users (e.g., accent, intonation) and culturally unique phrases, slang, etc. Or, it might omit culturally unique markers.	(1) Reduced pay rates or even job replacement for content creators. (2) Those without access to generative speech tools may experience competitive disadvantages or become relatively less productive than those with access to such tools	(1) Speakers experiencing adverse sentiment when their speech is altered or mistranslated by AI. (2) When AI errs, it takes more effort to correct AI	(1) Unknown impact when speech support becomes available at large scales. (2) Individuals with disabilities but without access to such tools might become even more marginalized. (3) Privacy leak through generative speech content	(1) Generative speech delivers wrong information. (2) Lower diversity in culturally and/or linguistically unique approaches to speech. (3) Causing threats to the unique culture of communities with disabilities. 4) Impact on existing physical and mental healthcare service systems to support individuals with disabilities. 5) Compared to processing other forms of data (e.g., text, visual), processing speech and audio data demands more computational resources.
Create new content: Users creating new voice and/or speech content using generative speech applications • Users producing long-form monologue or narration from text • Users creating content that impersonates a famous person/character • Users creating interactive, multimodal art installation • Users generating dialogues with famous people to share on social (e.g., make Rihanna say that she's my best friend) • Users creating virtual characters (e.g., for anime) • Users creating and editing podcast content	(1) Stereotyping or removing culturally unique topics and content. (2) Overly emphasizing certain speech/voice characteristics in others' speech without consent. (3) Misrepresenting certain speech/voice characteristics in others' speech without their consent. 4) Ignorance of personal preferences and needs for support.	Job replacement for content creators.	(1) Need for extra resources to check the quality of work produced through generative speech. (2) When AI errs, it takes more effort to correct AI.	(1) Unknown impact on creator-audience relationships when creators' content is not all produced by themselves. (2) Additional techno-stress for content creators to adopt generative speech tools to increase productivity. (3) Privacy leak through generative speech content.	(1) Generative speech content delivers misinformation. (2) Lower diversity in creative content. (3) Causing threats to the unique culture of communities with disabilities. 4) Impact on labor/job market for content creators. 5) Compared to processing other forms of data (e.g., text, visual), processing speech and audio data demands more computational resources.
Add speech to content: Users adding voice and/or speech content to existing, non-audio content • Users re-dubbing a video in a foreign language with the original actor's voice • Users re-dubbing a video in their own voices • Authors producing audiobooks in their own voices and/or in different languages	(1) Overly emphasizing certain speech/voice characteristics in others' speech without consent. (2) Misrepresenting certain speech/voice characteristics in others' speech without consent.	Reduced pay rates and even job replacement for certain content creators (e.g., video editors and narrators).	Need for extra resources to check the quality of work produced through generative speech.	(1) Unknown impact on creator-audience relationships when creators' content is not all produced by themselves. (2) Additional techno-stress for content creators to adopt generative speech tools to increase productivity. (3) Privacy leak through generative speech content.	(1) Lower diversity in creative content. (2) Compared to processing other forms of data (e.g., text, visual), processing speech and audio data demands more computational resources.